ASSESSMENT OF INHERENT AND OPERATIONAL ERRORS IN GEOGRAPHIC INFORMATION SYSTEMS

By

DALE ROBERT LIGHTFOOT

Bachelor of Science

in Industrial Arts Education

Oklahoma State University

Stillwater, Oklahoma

1980

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
July, 1986

ASSESSMENT OF INHERENT AND OPERATIONAL

ERRORS IN GEOGRAPHIC INFORMATION

SYSTEMS

Thesis Approved:

Stephen J. Walsh

Thesis Adviser

David R. Butler

John D. Vitek

Norman N. Durham

Dean of the Graduate College

PREFACE

This thesis was conducted as a case study for examining the amount of inherent, operational, and total error possible in products of a typical geographic information system. The error inherent in commonly used base maps or GIS data layers was assessed by comparing map data to field data at specific sample points located within the study area east of Stillwater, Oklahoma. GIS data layers used by this study were: (1) landcover; (2) slope angle; (3) slope aspect; and (4) soil type. Operational error is a result of human error and computer algorithm error created by the GIS process, while total error is a combination of both inherent and operational error and relates to the actual accuracy of any GIS product. This study calculates operational error and the theoretical minimum, maximum, and actual total error levels that result from various combinations of the four GIS data layers described above.

I wish to gratefully acknowledge the help given to me by my major adviser, Dr. Stephen J. Walsh. His ideas and guidance helped create this thesis while his assistance and understanding allowed it to be completed. Appreciation is also expressed to the other committee members, Dr. John Vitek and Dr. David Butler, for their invaluable comments and editorial assistance during the drafting of the manuscript.

Thanks are also extended to Mark Gregory of the Center for Applications of Remote Sensing at Oklahoma State University for his advice and assistance during the creation of the digital map products used in this study, and to Gayle Maxwell of the Cartography Service at OSU for her advice and help in preparing the maps and figures included in this thesis.

Special thanks go to Tim Loesch, a fellow graduate student, who volunteered his time and spent several cold, muddy days with me in the field carrying survey equipment and locating sample points. I wish to thank Jim Henley, a soil scientist with the Perry, Oklahoma Soil Conservation Service office who spent many hours donating physical work and expert soils analysis in the field to allow the soils portion of the fieldwork to be accurately accomplished. I appreciate the interest in this project shown by Billy Wagner of the Oklahoma State SCS office and for his help in arranging for Jim Henley to assist me.

Much appreciation is extended to Dr. Brian Carter of the OSU Agronomy department for timely soils information and materials, to the Geography department of OSU and Dr. Stephen Walsh for the funding they provided me during my studies, and to all of my friends in the graduate student warren who offered advice, condolences, and good conversation throughout the semesters.

Finally, appreciation and much love go to my wife, Tory, for continuing to provide much needed comfort and companionship, to my mother who stimulated many boyhood interests and initially convinced me to attend college, and to my father who was not able to see me finish my studies.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Before computers came into common use, spatial information was passed to users in the form of manually produced maps, charts, drawings and text. As the use of computers has grown, manual production methods have progressively been replaced by digital ones, and computer drawn maps are now common (Lehan, 1986). A computer generated map, manipulated by a geographic information system (GIS), is a powerful tool for summarizing and presenting complex spatial information. As useful as such information is for resource evaluation and appraisal, it is possible that products resulting from such analyses are in considerable error. The types and sources of such error are described by Vitek, Walsh, and Gregory (1984). The questions raised within their paper are the focus of this research.

The objective of this study is to evaluate the accuracy of geographic information systems. Errors inherent in layers of data on thematic maps used within a GIS often create final products of dubious quality. Furthermore, as the number of data layers increase, the combined error present in a GIS product also increases. The specific hypotheses to be tested are: (1) Inherent error exists within geographic information systems. Landcover, soils, and terrain data secured from Landsat satellites, Soil Conservation Service county soil surveys, and U.S. Geological Survey digital terrain tapes, respectively, contain error when compared to ground control data; (2) Operational error exists in GIS because of the inherent error of the source data. This error is manifested through data overlay and other data manipulation techniques. Total or composite error, which results from combining both inherent and operational error through the process of stacking data layers, increases as the number of layers used within the GIS increase. The null hypothesis is: data planes utilized in a GIS process contain no inherent errors and contribute to an error free final GIS product.

Nature of Geographic Information Systems

The inventory and monitoring of land use changes is generally considered basic to almost any resource management, planning, or land related program - either rural or urban, or local, regional, or national in scope. Society's growing awareness of a changing environment and its consequences has generated a need to know the land uses and activities present on the landscape and where, why, and how quickly shifts in these land activities are occurring (Henderson,1980). For over twenty years, geographic information systems have been growing in stature as more firms and governmental agencies realize the need for automated methods to retrieve, analyze, and display geographic data. This has been paralleled by a growth in the number of information systems available and consultants to develop and manage them (Crosley, 1985). Operational applications of GIS today include such areas as land and resource management, traffic planning, marketing, military planning, and a wide variety of other uses (Marble and Peuquet, 1983).

Numerous public agencies are currently developing and utilizing geographic data. Such agencies include: U.S. Census Bureau, U.S. Geological Survey, Central Intelligence Agency, National Aeronautics and Space Administration, and the Soil Conservation Service. In addition, an increasing number of commercial organizations generate geographically referenced data (Teicholz, 1980). These resource departments are faced with increasing environmental complexities in resource management decision making: they must consider earth sciences, natural resources, biologic, and socioeconomic information to identify development and policy alternatives (Guptill, 1981). These requirements for increased planning in land resource management have generated a need for greater quantities and varieties of basic resource information. This need is now being met by sophisticated data acquisition, processing, archiving, and retrieval methods.

Foremost among these methods are computerized geographic information systems (GIS) in which land resource data, such as soils, terrain, and landcover information are stored as geographically referenced layers of spatial information. The overlaying of one variable with another within the GIS allows for the two separate variables to be analyzed together in order to determine their combined, multiplicative effects and to observe the spatial pattern of variable interactions.

Although remote sensing derived data are used as inputs to GIS, the most common data source has been the analog map and in nearly all cases the input phase of GIS is heavily or entirely oriented toward creation of digital files from map documents (Marble and Peuquet, 1983). One common approach to map overlay (manual GIS) is to use transparent media (e.g. mylar or acetate) for mapping in order that maps may be superimposed for examination of the spatial consequences of variable interactions. This method, however, is cumbersome and inefficient for examining numerous variables. In addition, the map reader must visually compare the maps and store a mental image of the interrelationships between variables (Chang, 1982). An automated GIS process, however, facilitates the recording of information with speed and repetition. The storage and data handling capabilities of associated computer systems, along with the potential for large areas of simultaneous coverage, further demonstrates the usefulness of an automated GIS system. Data regarding environmental parameters can be processed into information relevant in a timely fashion for management decisions (Estes, 1982).

Analysis of the environment is sometimes limited by lack of quality data and the incompatibility of data derived from several sources (Vitek ,Walsh,and Gregory, 1984). A major problem affecting the use of available digital spatial data is this inability to combine different types and sources of data into a common data base (Teicholz, 1980). The development and implementation of a GIS can successfully lessen data integration problems and the time consuming process of synthesizing large amounts of information for problem analysis (Vitek, Walsh, and Gregory, 1984).

Output products from a geographic information system may be in the form of maps, tabular listings, temporary display on a computer CRT screen, or data permanently stored in computer files. These output products may be conveniently manipulated and analyzed to develop effective management plans for small or large areas.

GIS products are used for invaluable assistance to understanding and managing resource problems. The primary limitation in the use of a GIS for problem analysis is the question of the accuracy of the input data. Errors in the input data may occur as locational (spatial coordinates) and non-locational elements (landcover at a site). Remotely sensed data used in GIS products for landcover assessment are gathered by non-contact methods of data collection based on spectral radiance. Accuracies of such products can be relatively low as compared to on-site methods of data collection (Estes, 1982). Any map or product produced through a geographic information system is a collection of input information which is a generalization of reality. As such, all maps or related products must contain some error. Any divergence between a map and the earth's surface is a distortion of reality and is termed "error". Unfortunately, individuals involved in the utilization of GIS for land management too often use the map products and tabular results from the GIS without being aware of the errors that are generated through the process (Figure 1). The spatial display and interaction of data facilitated through geographic information systems can only be assured if the accuracy of the final map is known (Vitek, Walsh, and Gregory, 1984).

Assessing Map Error

Maps generally instill a level of confidence to the user that may not always be substantiated during analyses. Maps are symbols of reality, and thus contain a variety of errors depending upon intended use, design specifications, and level of generalization. The study of map error is not entirely new to cartography. Cartometry has a long tradition as a cartographic subdiscipline, but

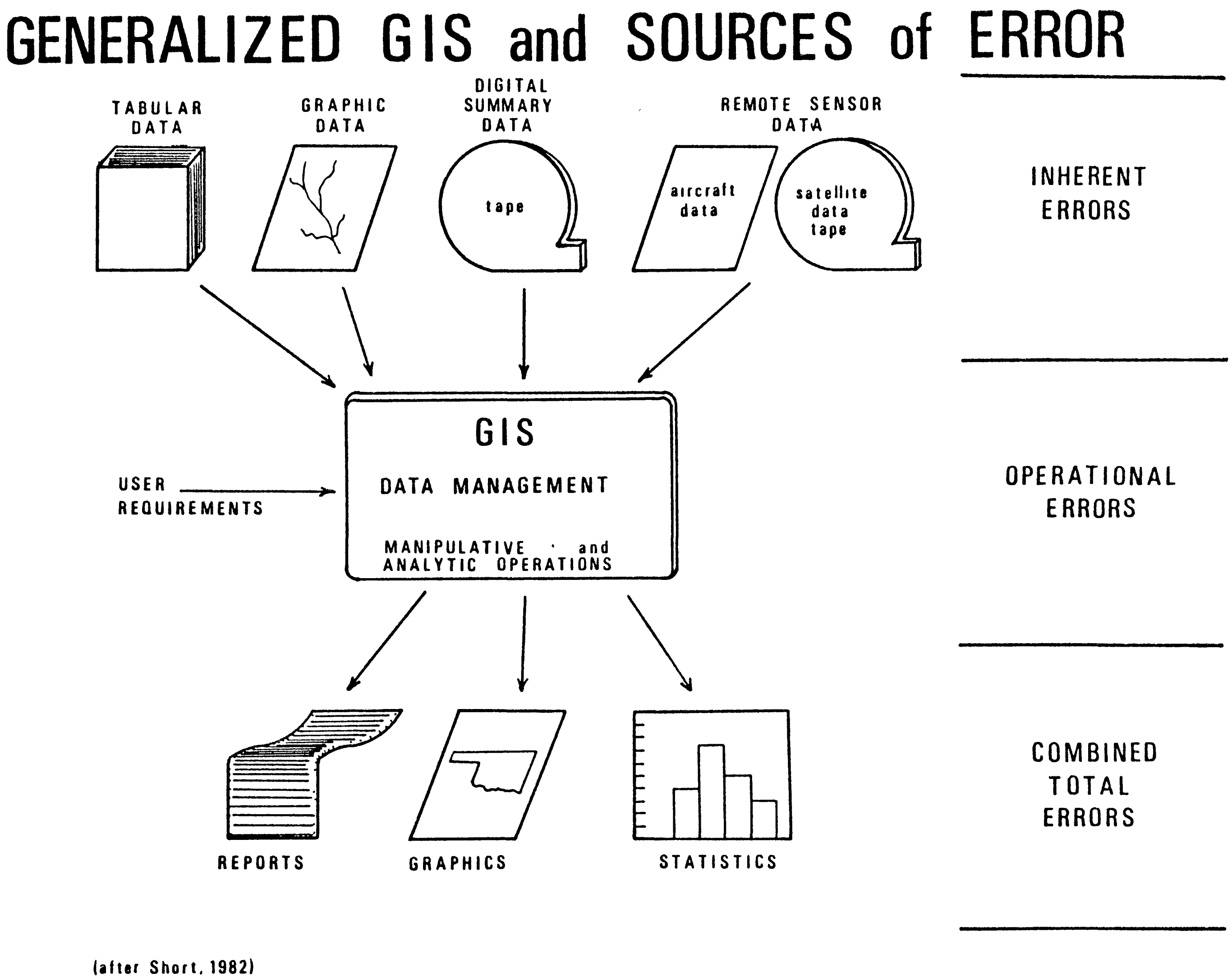


Figure 1. A Generalized GIS and Sources of Error

developments have been slow (Chrisman, 1982a). Computer cartographic techniques, being relatively new, have been scrutinized less than traditional cartographic methodology and errors produced by GIS operations, being newer still, have been only theoretically alluded to.

Geographic information systems are used to produce inventories of resources in spatial units and to produce maps or other products specifying the location of these resources. These products must be of known and reasonable accuracy to be acceptable to any decision process (Wehde,1982). To determine the accuracy of any GIS product, the possible sources of error in the GIS process must be addressed. Errors may be either inherent in the products of data capture or produced through data manipulation operations during the GIS process. Both types of error combine to contribute to a reduction in the accuracy of products generated by a GIS (Vitek, Walsh, and Gregory, 1984).

Inherent error deals with the quality of the data that is initially input into the GIS analysis process. Base maps or charts, tabular listings, digital data sets, interpretations of remotely sensed data, and other items comprise the individual data layers or overlays within the GIS system. The presence of error on maps is a basic tenet within the field of cartography. Attempts can and should be made to reduce error on maps but error cannot be completed eliminated (Vitek, Walsh, and Gregory, 1984).

Sources of inherent error include: (1) process of projection and transformation used to make the map; (2) symbolization scheme used to portray data and transmit information within a map; (3) scale of the map; (4) particular assumptions and methods employed; and (5) any other process which results in the generalization of reality within the map product. Despite the care in selecting, designing, and constructing maps, inherent error will always be present and is only increased through data manipulation proceedures (operational error) within the GIS (Vitek, Walsh, and Gregory, 1984).

Operational error increases the total amount of error based on the premise that inherent error cannot be eliminated but only enhanced by operational procedures. Operational errors are categorized as manipulative errors, data extraction errors, and data comparison errors. They are generally influenced by factors such as: (1) class interval selection; (2) use of various point, line, and area data from different base scales for combined overlays; (3) selection of polygons or various sizes of grids during data extraction; (4) interpolation of data errors, and (5) errors in the digital alignment of the various layers of geographically referenced data (Vitek, Walsh, and Gregory, 1984). Such digital alignment or positional errors are of critical importance to many users because of the problems associated with error accumulation due to overlay operations (Marble and Peuquet, 1983).

Study Objectives

In order to demonstrate the level of accuracy of any final product, accuracy statements should be included as a part of the GIS output product; without such error designations the map user may draw false or misleading conclusions. Methods of specifying map accuracy should be aimed at map users with a minimum of statistical training in order that the assessments be more widely used. Standardization of accuracy tests should enhance estimations of map accuracy (Vitek, Walsh, and Gregory, 1984). This study will develop a methodology to test for map accuracy and will also incorporate standardized map accuracy statements concerning final GIS output products.

Previous GIS error related studies have focused only on specific operational errors concerning the testing of methods of data capture (polygons or various cell size, shape, and/or location) (Wehde, 1982; Henderson, 1980), or they have focused on methods of data manipulation by the GIS (Aronoff, 1982a and 1982b; Estes 1982; and Newcomer and Szajgin 1984). These studies have rarely alluded to potential problems concerning errors inherent in base products used for any GIS overlay procedure. This study considers these inherent errors as equally important to the overall accuracy of GIS output products and, therefore, will address actual inherent errors as its starting point and will follow these inherent errors through the operational, data manipulative process. After the accuracies of each base product have been assessed, this study will calculate the overall error within a finished product based on the combined multiplicative effects of inherent errors and operational errors. This study will employ a quantitative analysis of all inherent and operational errors. The methodology will include the sampling of selected variables in order to initially assess base product accuracy. The GIS base product variables to be examined are soil type, landcover, and slope angle and slope aspect. These four variables were chosen because of their common use in GIS applications.

Study Area

The study area for this analysis is a four mile square (10.4 square kilometer) area located east of the Lake Carl Blackwell dam in Payne County, Oklahoma. This site covers a region having an adequate diversity of soils, landcover, and a gently rolling surface terrain (Figure 2). The selected four variables are judged sufficient to include the more important variables utilized in most land use analyses involving GIS. Landcover, soils, and terrain variables are customarily combined in order to create a dasymetric land use map. Dasymetric mapping is a type of statistical or quantitative mapping which simultaneously examines two or more maps by conceptually or physically superimposing quantitative data (Chang, 1982).

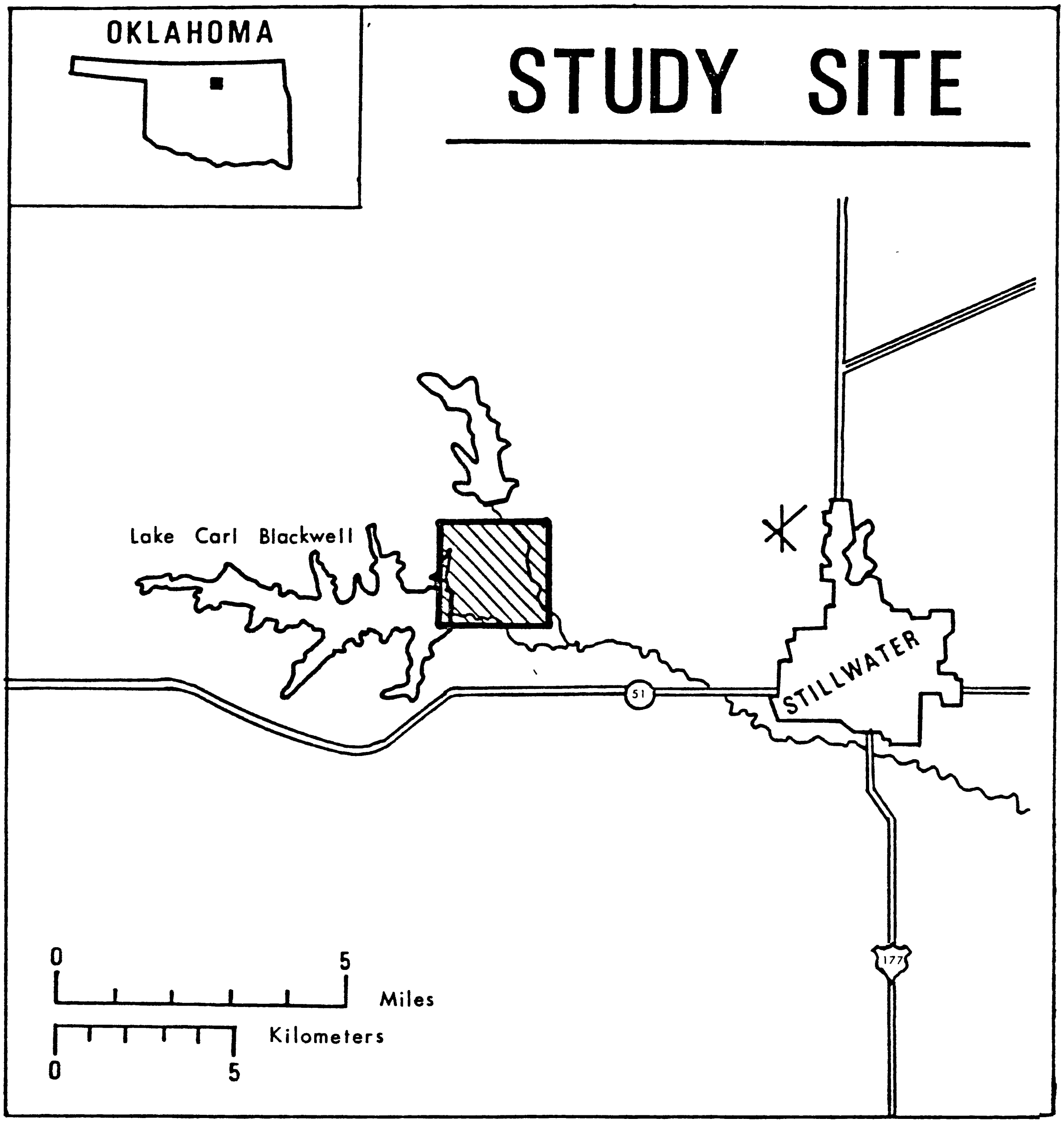


Figure 2. Field Study Site

A more rugged topography would likely introduce more error into the analysis. The results of this study should serve as a useful baseline of average error to be expected from a GIS evaluation process. Because generalized map data aggregated by selected cell sizes will be compared to point sample data collected in the field, more error than might normally be expected in thematic map layers or GIS products may result. As part of the study of inherent and operational errors, two different grid sizes will be evaluated as to their impact on GIS accuracy. The grid sizes used will be 2.5 acre (100 square meter) and 10.0 acre (200 square meter). Two methods of cell data characterization (center point of cell and cell dominant) will also be compared.

Thematic Maps

A brief discussion regarding the method of data collection for the construction of thematic maps used in this analysis is necessary to explain sources of potential inherent error within the data. The thematic data utilized in this analysis, soils, landcover, and terrain, are described.

Landcover Maps

A common type of landcover map is produced by processing Landsat digital tapes. A generalized scheme for creating a landcover map from an unprocessed Landsat data tape would include: (1) preprocessing the Landsat data to remove unnecessary banding or striping and reformatting the data into a more efficient format; (2) selection of study area boundary coordinates to reduce the size of the area being processed and thereby minimize processing time; (3) creation of a set of spectral categories using a search routine to develop the basis for a maximum likelihood classifier; (4) running the classification program for the study area selected, resulting in a thematic classification of the landcover as recognized by Landsat; (5) fine tuning this classification by combining spectral classes as necessary; and (6) geographically referencing the Landsat thematic classification to the Universal Transverse Mercator coordinate system (Walsh, 1985). The result is a landcover map which is, ideally, both geometrically and spatially correct. An attempt is always made to keep georeferencing error to one-half pixel (picture element) or less discrepancy between Landsat data and the map used for georeferencing.

Terrain Tapes

The digital terrain tapes, used for the creation of both the slope angle and slope aspect maps, are produced by the U.S. Geological Survey from maps at a scale of 1:250,000. Terrain data on the USGS 1:250,000 scale topographic maps are first separated into two types of data: (1) elevations as contour lines and points and (2) stream and ridge lines. These data are then sorted, matched, and resorted to obtain a digitized grid of elevation values for every 0.01 inch (0.25 mm) on each map (about 200 feet or 60 meters on the ground). Undefined points on the grid are found by either linear or planar interpolation (USGS, 1979). Additional related products created by the USGS are the digital elevation models (DEM's). These digital products also depict terrain but differ from the digital terrain tapes primarily by the sampling methodology used to derive the product. DEM's produced at a 1:250,000 scale are created by sampling elevation data from contours and ridgelines on the 1:250,000 scale topographic maps at intervals of three arc seconds, which represent about 295 foot (90 meter) intervals in the north-south axis and a variable dimension of 295 feet to 195 feet (60 meters), from the equator to 50 degrees latitude, respectively, in the east-west axis because of convergence of the meridians. As with the digital terrain tapes, these sampled values are then digitized to create an elevational grid. DEM's produced at the 1:24,000 scale are likewise created by sampling a 1:24,000 scale (7.5 minute series) topographic map at intervals of 100 feet (30 meters). The 1:24,000 scale DEM's are stored in one of two separate DEM data bases depending on the tested vertical accuracy of each product. Accuracy is either less than 7 meters vertical root mean square error (RMSE) or 7 to 15 RMSE. The accuracy of all digital terrain tapes and digital elevation models is highly dependent on the vertical accuracy of the original topographic maps from which they are created (Elassal and Caruso, 1983).

Soils Maps

County soil survey maps, as used in this study, are produced by Soil Conservation Service soil scientists by combining information on the landscape with prior general knowledge of the soils within the region. These data are then used to establish initial boundaries between soil mapping units. Through the use of both a hydraulic soil probe and a hand soil auger, soil cores may be taken and classified throughout an assumed soil mapping unit to more accurately define the suspected boundaries of these soils. Use of ancillary data sources such as color infrared and stereoscopic aerial photographs are also incorporated into the analysis to help confirm the boundaries of the soil mapping units. Sampling intervals between soil cores taken in the field will vary considerably with the complexity and diversity of the soil; a greater number of core samples are taken in areas with more complex soils. Sampling strategy may allow samples to be taken at regular intervals along transects. Cores are often taken at irregularly spaced intervals and random locations depending on any physical barriers encountered and/or the mapping bias of a soil scientist regarding criteria used for delineating boundaries between mapping units (Henley, 1985). Soils found within a mapping unit that exhibit different physical properties, such as texture, structure, color, and which therefore constitute a different soil type, are termed soil inclusions. Soil inclusions within any soil mapping unit may vary in both size and area occupied depending on the scale at which the soil map is produced. At the 1:20,000 scale county soil surveys, soil inclusions as large as 4 acres (1.62 hectares) in size and occupying as much as 20 percent of the total area of any soil association will not be mapped. Similar sized inclusions occupying up to 25 percent of the area of any soil complex are likewise not mapped at a 1:20,000 scale map (Henley, 1985). The possibility of a particular soil inclusion being found within any mapping unit is, however, acknowledged in the text of the soil survey. The accuracy of any SCS soils map, then, will vary depending on: (1) complexity of the soils; (2) experience of the mapper; (3) availability and utilization of ancillary data sources; and (4) number, size, and extent of soil inclusions (Henley, 1985).

Chapter II discusses the literature important in the accuracy assessment of geographic information systems. Additional references concerning the accuracy of data layers to be used in the GIS are also discussed. Chapter III discusses the procedures employed during fieldwork and the creation of thematic maps. The methodology used to quantify the accuracy of GIS products is also presented in Chapter III. Chapter IV presents the results of the analysis, while Chapter V offers conclusions to the study along with supporting evidence.

CHAPTER II

REVIEW OF THE LITERATURE

The sections that follow will group all of the literature related to this thesis into similar topical categories for discussion. The following discussion describes the major focus of each article and, where appropriate, its relationship to this paper.

Geographic Information Systems

A vast amount of current literature exists which describes the hardware, software, and operational considerations of remotely sensed imagery and GIS. References described which report on the operations, uses, benefits, and data handling/interpretation problems within GIS serve as the organizational topics for the review of the literature.

The initial development of GIS focused on designing computer software for the production of analytical output products. Manipulation of data is increasing in complexity due to increases in the speed and power of data storage and computer manipulation. Considerable effort still is required, therefore, to complete the development of GIS

(Vitek, Walsh, and Gregory, 1984). Automated geographic information systems have many advantages in comparison to manual methods of using geographic data. They save money, time, and other resources; they provide better products in a more timely manner; they provide powerful tools of analysis and clear, attractive data displays. There are many world trends leading to the increased use of GIS: increases in the needs of the world's people, increases in the quality and quantity of technology on which GIS depend, and increased awareness of the value of GIS systems. Automated GIS could be applied to many world problems, but that would first require the creation of very large or even global data bases (NASA and ESRI, 1984).

The GIS is distinguished from other information systems (e.g. management information systems for business applications) by its focus on spatial units and area relationships. For a user who is faced with developing a set of baseline maps and data pertaining to land use and landcover, updating such maps and data periodically, and relating several sets of associated data to information about land use, the use of a national land data system with a geographical computer capability is an invaluable asset. In the past, a GIS was often designed to meet only the needs of a specific problem and the data capture, management, and analysis functions were restricted to the unique charateristics of specific data sets. More recently, systems are being designed for generic data types and functions and provide much greater flexibility and a wide range of applications which could allow for a large spatial data base, such as the national land data system (Anderson and Marx, 1985). There are a number of significant and unresolved technical problems in GIS operation that limit both the size and efficiency of current systems. The improper design of a GIS is the main cause of system failure (Marble and Peuquet, 1983). At present, the interface between GIS and remote sensing is weaker than it should be, and each side suffers from a lack of critical support which could be provided by the other. The GIS has a need for timely, accurate updates of the various spatial data elements held in its system and remote sensing systems could benefit from access to highly accurate ancillary ground data which could improve classification accuracies (Marble and Peuquet, 1983). A GIS requires a method which must handle quality components along with the data directly depicted on a map. Quality information includes lineage records, accuracies of position and classification, integrity of data structure, and temporal reference. This quality component informs users of the suitability of data for their applications (Chrisman, 1984). Determination of a boundary line to represent the absolute limits of a geographic distribution of dispersed entities is often non-trivial, yet the absence of conventional algorithms and methods of boundary definition has often relegated the process to an exercise in "eyeballing". The problem can be conceptualized as one of trying to estimate the true but unknown location of a line which has some degree of existence in reality, but in other cases no real line can be said to exist. Unfortunately, the latter type of problem seems to be much more common than the first (Averack, 1984).

Natural resource analysts have traditionally focused on the use of field methods that rely on the professional judgement of on-site specialists. With the advent of more extensive land developments affecting larger geographic areas, natural resource analysts are turning to maps, aerial photography, remote sensing, and other methods in an attempt to describe area-wide field relationships with some known level of accuracy (Salmen, 1978). The recent rapid development of remote sensing into a seperate field of study has left cartographers uncertain of the role remote sensing plays in their future, especially because many remote sensing specialists assert that their field is a new discipline independent of cartography. The principle difference between the two fields lies more in the technology each field applies in processing environmental information than in their respective objectives. The fields of cartography and remote sensing are essentially similar from the practical as well as the theoretical viewpoint. Both disciplines are concerned with processing environmental information so the environmental processing needs of the future could best be met by combining the current activities of cartography and remote sensing into a single improved strategy for handling geographic information (Kimerling, 1976).

Estes (1982) states that in order to realize a greater potential for effective resource management, both the philosophy and conceptual linkages between remote sensing and GIS need to be improved. Estes further suggests that additional directions for future research in GIS are to link artificial intelligence systems and GIS, and to create interchangeable techniques for both software and hardware for all such related systems. Estes (1985) concludes that considerable research and technique development are still needed if we are to increase the geographic potential of remotely sensed data. To enhance our ability to map, monitor, and model a variety of environmental conditions and processes, advances in several relevant research areas must be incorporated into a systems approach to the processing of remotely sensed data. Key technologies to be integrated include geographic information systems, and elements derived from the field of artificial intelligence. A GIS may be the key to effective use of these combined data in many geographic applications (Estes, 1985). The integration of techniques from spatial database, artificial intelligence, and image processing show significant promise in overcoming several of the major obstacles preventing GIS from handling large, heterogeneous spatial databases in an efficient and flexible manner (Peuquet, 1984).

Starr (1982) reports that the creation of an integrated digital cartographic data base, to contain data from various sources and to provide a variety of standard and customized map and digital data products, is essential for effective resource evaluation. Natural resource managers face increasing complexities in their decision making. They must consider earth sciences, natural resources, and biologic and socioeconomic information to identify development alternatives. Systematic procedures for evaluating alternatives must be devised to assist in reaching sound conclusions. These procedures often require rapid manipulation and flexible use of spatial data from a variety of sources (Guptill, 1981). There is a rapidly growing need to use GIS to manage extremely large databases containing data integrated from a number of imagery, cartographic, and other sources. Current GIS technology is, however, exhibiting severe shortcomings in meeting these performance demands because geographic data possess a number of characteristics not found in other types of two or three-dimensional data: (1) geographic boundaries tend to be convoluted and irregular; (2) data in digital form tend to be incomplete, imprecise, and error-prone; and (3) spatial relationships tend to be vague or application-specific. Current demands on GIS technology, then, require major advances in models to represent geographic phenomena more efficiently, in flexible procedures for searching complex geographic databases, and in developing methods for dealing with imprecision (Peuquet, 1984).

The impact of soils, landcover, terrain, and climate on non-point pollution can be assessed through the data overlay analysis framework afforded through geographic information systems (Walsh, 1985). GIS techniques facilitate the assessment of spatial interactions of multiple variables, whether through manual or automated techniques (Walsh, 1985). Automated scanning (laser line following) digitization may be up to five times more accurate than manual digitizing. The automated laser scanning device is able to keep virtually all of its measurements in a band only two scan line widths apart. The average band width deviation created by a manual operator is ten line widths across (Chrisman, 1982a). While human involvement has advantages for corect data assessment, it nevertheless causes major problems with respect to high running costs, large time consumption, and many errors being generated owing to the presence of a human operator (Marble and Peuquet, 1983). Three fundamental distinctions between digital representation of cartographic data and the conventional printed map are: (1) the need, in digital work, to explicitly encode the spatial relationships among various elements of the data; (2) the need, in digital representation, to numerically encode attributes of the cartographic data that are normally conveyed to the reader of a printed map through color, line weight, symbology, and labels; and (3) the fact that digital data are irrevocably bound into a computer environment that was not developed with spatial data handling in mind (McEwen, Calkins, and Ramey, 1983).

Analysts interested in using digital computers for the analysis of natural resource information have turned to grid cell data structures. A grid cell converts the original form of thematic data to more digitally compatible information and controls location through the imposition of an arbitrary rectangular coordinate grid overlaid on the maps. This superimposition of the grid on the mapped thematic format adds a second level of generalization to the already generalized original map source. Most of the criticisms about the use of grid cells in natural resource analyses are directed at this application of a second level of generalization to a data source already inherently generalized (Sinton, 1978). The cell unit seems to be of high relevance for incorporation into a complete and integrated GIS. The cell can be used as the common areal unit for many different socio-economic and physical phenomena of geographic concern. The cell defines the lowest resolution of both value and location, and a truly spatial study must take the cell (and cell arrangements) as the more appropriate unit to deal with than the a-spatial individual of traditional statistics

(Wallin, 1984). The number of polygons created in an overlay depends not on the number of polygons being overlaid, but on the complexity of each polygon, defined by the vertices. Moderate numbers of polygons are produced when the overlaid maps show statistical independence, meaning arcs on one map show no tendency to follow arcs on another map. When arcs show a tendency to coincide, however, as when a prominent linear feature (such as a road) appears in the image of many different maps, spurious polygon problems arise. The presence of spurious polygons in an image resulting from overlay presents major complications because the image is made more complex than necessary and the volumn of the polygon dataset can multiply by many times (Goodchild, 1978).

Teicholz (1980) suggests that a major problem facing planners, resource and marketing analysts, quantitative geographers, and others is the ability to combine different coverages of data (population, land use, sales areas, zoning districts) into a common data base, and the ability to compare these irregular coverages. The problems of aligning data from differing geographic bases and various data collection sources have plagued researchers trying to apply these data to work in their respective fields. Increasing numbers of local, state, and federal agencies collect all types of aggregate data and report the data in all types of formats. This lack of conformity between data sources has hindered the statistical comparisons of data sets collected from different geographical bases (Matson, 1985). Gates and Heil (1980) report that the major criteria for examining spatial information systems are dependent on: (1) intelligent data; (2) topological character; and (3) adherence to general purpose coordinate systems. If any spatial information system has all three attributes it may be useful for most GIS demands. The absence of one or more of the three attributes reduces the potential capabilities of the system (Gates and Heil, 1980).

Sources of Error in Geographic Information Systems

The major focus of this thesis follows questions raised by Vitek, Walsh, and Gregory (1984), and it is in response to questions generated by that paper that this thesis was designed. The two major causes for GIS output errors are: inherent errors in the base products used for input into geographic information systems, and operational errors caused by data integration within the geographic information process (Vitek, Walsh, and Gregory, 1984). Every map contains inherent error because of the nature of map projection, construction techniques, and the symbolization of data. Operational error is introduced through data entry, data manipulation, data extraction, and data comparison within GIS (Vitek, Walsh, and Gregory, 1984). The result can be poorly presented maps and geographic material which fail to impart the information intended or, even worse, mislead (Robinson and Jackson, 1985). Because a map user may draw false or invalid conclusions from inaccurate maps, any person or agency responsible for the creation of GIS products must be responsible for the specification of the amount and type of error in the product (Vitek, Walsh, and Gregory, 1984). When presented with a set of data, users should attempt to understand the level of generalization that has taken place on the original observations and measurements because the procedures used to generalize or abstract data will significantly affect the utility of such data for analytic purposes. The critical issue is the extent and nature of the detail lost in the process of generalization, and the effect that this lost detail has on the thematic content of information. Data for which no record of precision and reliability exists should therefore be suspect because once data is entered into a computer it assumes an aura of respectability (Sinton, 1978). A renewed emphasis is needed concerning techniques for producing thematic products of known accuracies. This will require new sensor processing systems and analysis techniques (Estes, 1985).

Marble and Peuquet (1983) report that the accuracy of a GIS product is dependent on characteristics inherent in the source products and user requirements. They also suggest that positional accuracies in the GIS, which are of critical importance to many users because of overlay analysis problems, can never exceed the accuracies of the original data source.

Mead (1982) presents a system for rating GIS products (on a 100 to -60 scale) based on the evaluation of certain aspects of the base products. He states that factors such as age of the data, areal coverage, map scale, map resolution, positional accuracies, content accuracies, and data format should all be examined before any final rating is given. A questionnaire devised by Mead and administered to state level GIS managers indicates how such experts view data quality problems in GIS systems and products. Results of this questionnaire indicate that all of those surveyed believe it is possible to measure data quality and 63 percent acknowledged data quality problems in their systems, yet only 53 percent ever include statements regarding data quality with their products. Local registration is the key to accurate positioning. Control points are established within a small area on a map by scaling and overlaying projected Landsat images to select features that have remained constant since the map was made. Many linear features can be plotted as accurately from Landsat Multispectral Scanner images as from aerial photographs, with an average relative error of about 100 feet (30 meters) when plotted on maps at a scale of 1:50,000. Non-linear features with an area of less than

2.5 acres (1 hectare), or two Landsat pixels, will not be positioned as accurately as linear features and may not be recognized at all (Gregory and Moore, 1986).

The results of a cartographic inventory or modeling effort are only as accurate as the data that were used (Hansen, Dwyer, and Mogg, 1985). Newcomer and Szajgin (1984) show how the highest accuracy of any GIS output product can only be as accurate as the least accurate data plane of information that goes into the process. The final product is less accurate than any of the individual layers utilized. The statistical formulae and concepts presented within this article serve as the means for arriving at overall product accuracy statements generated by this thesis. They detail procedures for determining statistical product accuracies based on the alignments of correctly or incorrectly labeled cells, and present a method to calculate both the theoretical upper and lower accuracy limits of any GIS product based on the levels of error inherently present in each individual layer of data.

One problem that makes map overlay and other computer cartographic operations so difficult is that of numerical errors. As database sizes grow, these initially trivial inaccuracies can cause topological inconsistancies that affect the overall integrity of the database. The usual solution is to represent the coordinates more accurately by using a finer grid. This can never lead to a complete solution, however, and is actually part of the problem

(Franklin, 1984). Jenks (1976) and Chang (1982) both discuss problems, errors, and ignorance concerning quantitative or statistical maps. Manufacturers of computer equipment focus the attention of geographers and cartographers on machine accuracy, implying that more precise equipment equates with more accurate maps. This implication is true to some extent, but many maps of inferior quality are still prepared with very accurate equipment. In many cases, the poor quality of published maps can be traced directly to human errors in digital aquisition. Human frailties, the physiological, psychological, and logical limitations of the geo-cartographer, are clearly evident when mapmakers display representations of naturally occuring lines created by computer driven plotters (Jenks, 1981). Jenks (1976) concludes that, although statistical maps are produced in greater numbers by persons in a wider range of disciplines, overall map quality has not improved. This lack of quality improvement is likely because of the lack of understanding of statistical map function, symbolic language of mapping, and cartographic data manipulation problems. This reference makes the determination that statistical maps are not good sources of specific information about places and that good maps can only be created by those who understand mapping processes. Chang (1982) discusses procedures for multi-component quantitative mapping. Because of recent developments in remote sensing and computer mapping, this method of simultaneously mapping two or more variables is becoming more useful and practical. Chang (1982) concludes by noting the urgent need to study communication effectiveness of multi-component maps from the standpoint of both the researcher and the general audience.

Map production policy and cartographic research into computer generated maps is discussed by Monmonier (1983). He relates algorithm and grid cell size accuracy to these digital maps. This reference also suggests that the digital cartographic data base will eventually replace the paper map as the primary medium for storing and analyzing geographic information. The reference is of interest to this thesis in its determination that accuracy in computer generated maps can almost always be improved by reducing the size of the grid cell. This topic will be further discussed in another section of this chapter.

Chrisman (1982a and 1982b) discusses why the processes used in digital map production are necessarily approximate, and then examines the potential for error utilizing an 'epsilon distance model' applied to geometric details on a study map. Although Chrisman's particular methodology will not be used in this study, general information concerning errors in digital data maps, used by GIS, is helpful to the purpose of this paper. Although error analysis cannot be fixed for all maps, a generalized discussion of all potential sources of error includes: (1)

location of ground position, which is hindered through errors in surveying and geodesy; (2) interpretation error, created by incorrect placement of boundaries or by misclassification; (3) scale of the map selected; (4) map projection error; (5) errors created during map drafting either through equipment problems or human operator error; (6) digital handling of data which introduces potential error through digitizing and rounding of figures by the machine algorithm; and (7) combined effects (Chrisman, 1982a). Only primitive GIS software support systems have been made available by manufacturers of manual digitizing tables, and many of the support systems available from other sources have been designed to deal with engineering drawings which have very different characteristics than maps. The individiual researcher, then, is faced with a situation in which he or she has no useful basis for the estimation of the time and resources required to successfully complete a data capture task, and only crude tools with which to undertake the operation. This often leads the researcher to the decision not to make use of modern technology for spatial data handling and this, in turn, contributes to decreased research efficiency because a major analytic tool has been discarded (Marble, Lauzon, and McGranaghan, 1984). The high cost for hardware and software restrict the application of automatic digitizing techniques to large projects that are voluminous enough to recover the investment costs (Wolf-Dieter, 1984).

Assessing the Accuracy of Geographic Information Systems

Inherent Error

A large number of sources exist on methods to determine the accuracy of land use maps produced by Landsat classification techniques. VanGenderen, Lock, and Vass (1977 and 1978) state the need to know the accuracy of any land use maps generated in order to achieve wider acceptance among users of land use mapping from remote sensing data. A GIS can even be used in the process of assessing the accuracies of thematic maps by: (1) selectively retrieving particular classes within the map; (2) compositing the selected data sets through spatial analysis techniques; and (3) determining the location and number of occurrences of various combinations of map classes, thereby determining where two or more data sets are logically mis-matched and the nature of the possible classification errors (Henderson,1985). Story and Congolton (1986) state that the most common method of expressing the accuracy of images or maps is by error matrices, or accuracy evaluation matrices, which calculate the percentage of the map area that has been correctly classified when compared with reference data or ground truth. These error matrices should always appear in the literature whenever accuracy is assessed so that the users can compute and interpret these values for themselves (Story and Congolton, 1986). In this kind of tally, the reference or field data are compared to the classified map data and the major diagonal across the matrix indicates the agreement between these two data sets. Overall accuracy for a particular classified image/map is then calculated by dividing the sum of the entries that form the major diagonal (the number of correct classifications) by the total number of samples taken (Story and Congolton, 1986). Rosenfield and Fitzpatrick-Lens (1986) argue that non-diagonal elements of the matrix have been neglected. Coefficients which utilize all cell values in the matrix, rather than only the correctly classified cells in the matrix diagonal, can be computed to correct for the chance agreement in classification which would inflate accuracy percentages. VanGenderen, Lock, and Vass (1977 and 1978), Story and Congolton (1986), Aronoff (1982a and 1982b), Ginevan (1979), Hay (1979), and Walsh, Vitek, and Gregory (1982) all describe specific methods for setting up error matrices, sampling designs, and statistical procedures for assessing the accuracies of digitally produced landcover maps. Techniques borrowed from these sources will also be adapted for use in the testing of soil and terrain map accuracies in order to check for total inherent errors within these various GIS data layers.

Operational Error

Possible sources of operational error are discussed in Vitek, Walsh, and Gregory (1984) as being manipulative (generalization and interpolation) errors, data extraction (search and measurement) errors, and data comparison errors. This last category deals with overlay (grid or polygon) problems, proximity (i.e. nearest neighbor statistic) problems, and contiguity and connectivity considerations. One of the major sources of operational error in GIS can be attributed to map registration which establishes the link between physical coordinates of the digitizing tablet and the geographic coordinates of map projection. The data capture stage of digitizing introduces the primary source of human error to the database. When manual digitizing is being performed, digital map quality is dependent on how accurately the operator traces map lines with the cursor. Polygon formation/verification and editing are additional sources of operational error because adjustments are again made to the data by the program and/or operator (Hansen, Dwyer, and Mogg, 1985). Digitizer operators cannot maintain line-following precision better than 0.004 inch (0.1 mm), yet the nature of this error has not been investigated by cartographers. In the past, line-following error has been treated as random "noise", yet line-following error is not a random phenomena but is systematical and related to the motor coordination abilities of the digitizer operator. An operator's knowledge of the error pattern, then, can often lead to decreased digitizing error (Traylor, 1979). Interactive graphic displays of the data during the digitizing and editing process, however, tend to reduce the production of error in GIS (Guptill, 1978).

Muller (1977) examines the effects of subdividing geographical territories into spatially ordered and gridded data for use with statistical maps and discusses the relationship between total map error and grid resolution/grid position. The transfer of data from irregular administrative units to a grid system introduces cartographic error and the mismatch of areas between the original map and the gridded map is related to both grid resolution and grid position. This reference also determined that total map error is linearly related to grid resolution and statistically independent of grid position. While a smaller grid size almost always provides greater accuracy for a digital map product, accuracies may also be improved by increasing the scale (thereby the resolution) of the digital map when working with a fixed grid cell size (Muller, 1977). Gersmehl and Gersmehl (1982) graphically demonstrate how the percent classification error of soils information will increase as the size of the grid cell used to generalize the actual distribution decreases.

For many applications a grid based data structure is preferable to a hierarchical polygon based data structure because it is easier to implement, update, and use. For the case of computerized GIS, grid based data structures have become widespread. The major advantage of grid base data is the ease of finding data items for a particular location. The accuracy of representing any given map pattern can be improved to any required level by decreasing the size of the grid cell (Crapper, 1984). Monmonier (1983), Wehde (1982), Henderson (1980), VanGenderen et al. (1978), Muller (1977), and Nichols (1975) also describe the effects of grid cell size in relation to product classification accuracy. All six sources determined that a smaller cell size usually resulted in increased map accuracies. Henderson (1980), however, found that smaller cell size did not always result in increased classification accuracies, although no consistent pattern of occurrence was found to control this phenomenon. A brief discussion of classification by cell midpoint and cell dominant area is given in Markham and Townshend (1981) although no comparative accuracy statements or assumptions are given. This reference also discusses the direct overlay method of grid map creation where grid cells are assigned the choropleth data value of the polygon into which they fall. This overlay method is the most commonly used in GIS because it has the advantage of being relatively simple to perform (Markham and Townshend, 1981).

Composite Error

The statistical procedures outlined in the previously mentioned Newcomer and Szajgin (1984) article serves as the basis for the assessment of total error expected from any GIS product based on the quantified accuracy of each of the three layers analyzed during inherent error assessment. This reference shows that composite error will increase dramatically as the number of layers used in the GIS increases or as less accurate thematic map data are utilized for base products. It is the only source found which quantifiably assesses composite error.

Base Products and GIS Data Layers

For general insight into the methods used to create the maps (GIS data planes) used during this analysis, several references were consulted. Wilson and Thomson (1982) present methods for the creation of digital files of Landsat, terrain, and other data and discuss how such files may be manipulated to produce useful output for planners and managers. Their discussions on the assembly of digital data files and the subsequent manipulation of such files provides a summary of how digital data layers are initially created. Variability of interpretation accuracy exists in landcover maps and this variation in accuracy is regionally distinct. The identification of such regional variation can serve to make remote sensing a more effective land use mapping tool by suggesting the best regional allocation of the most efficient data collection systems, which for only some places will be satellite imagery (Schwarz, 1976). The U.S. Geological Survey Professional Paper #1175 (1980) gives information on the national map accuracy standards applied to the creation of the topographic maps used for the terrain map information in this study. This reference states that 90 percent of the map features tested at scales of 1:20,000 or smaller must be accurate to within 1/50th of an inch (0.05cm) of the correct map position. Although such map testing is traditionally performed by field surveying on a sample basis, several photogrammetric methods, such as aerotriangulation of extra points, are used. The discipline of cartography is undergoing profound changes that center on the emerging influence of digital manipulation and analysis of data for the preparation of cartographic materials and for use in GIS. Operational requirements have led to the development by the USGS National Mapping Division of several documents that establish in-house digital cartographic standards. The documents have been assembled into the USGS circular #895, which consists of several chapters. The first chapter is a general overview of digital cartographic standards (McEwen, Calkins, and Ramey, 1983), and the second chapter describes the digital elevation models produced and distributed by USGS (Elassel and Caruso, 1983). Succeeding chapters are made up from documents that establish standards for other types of digital cartographic data currently produced.

The National Cartographic Information Center's publication on digital terrain tapes (1979) provides information on the digital terrain models frequently used by geographic information systems. This reference discusses the sampling procedure used to create digital terrain tapes. Two types of data: (1) elevations as contour lines and points; and (2) stream and ridge lines are sorted, matched, and resorted to obtain a grid of elevation values for every 0.01 inch (0.25 mm) on each 1:250,000 scale map. This distance represents approximately 200 feet (60 m) on the surface of the earth. This source points out that the accuracy of any digital terrain tape is only as good as the accuracy of the 1:250,000 scale topographic map from which it was created. A related source by Elassal and Caruso (1983) discusses the sampling scheme and the accuracy standards of similarly produced digital elevation models. The 1:24,000 scale DEM covers one 7.5 minute USGS topographic quadrangle and provides digital topographic data sampled at intervals consistent with the contour level of each of the original 1:24,000 scale maps. The 1:250,000 scale DEM is created using the same sampling scheme as the 1:24,000 DEM. A digital elevation model represents topographic or other surfaces in terms of sampled values, and includes a rule for regaining a continuous description of the surface. Estimation of the quality of this representation is important for planning the sampling and reconstruction and also for evaluating the accuracy of products derived from the digital elevation model. The accuracy of a digital elevation model depends on the size of the sampling interval in relation to the variability of the surface. The method of reconstruction, in comparison, is of less significance (Tempfli, 1982).

Anderson et al (1976) present the guidelines used for Landsat landcover classification. Standardized categories of landcover classification are presented. Classification levels, indicating increasing complexity of detail to be mapped, are I, II, III, and IV and are determined to some extent by the scale or resolution of the base product being utilized to determine landcover divisions. This thesis classifies landcover according to category I classes which, according to USGS, is the classification level best adapted for use with Landsat data. Nichols (1975) describes the accuracies of computer generated soils maps and provides data to compare detailed soils maps to gridded soil maps generalized through cell dominant data capture. As cartographic detail on the original soil map increased, more disagreement (error) was found between the original map and the computerized soil map. In addition, more error was found in digital soils maps created with larger grids than in those utilizing smaller grids (Nichols, 1975).

The SCS Soil Survey Manual (1984) includes detailed descriptions of the nature, purpose, and uses of soil surveys. Related topics such as mapping preparations, examination and description of soils, mapping units, mapping legends, and field operations are also described. While most of this information has few direct applications to the focus of this study, one section explains the accuracy limitations (Table VIII) imposed on soils mapping and acknowledges the level of generalization (and therefore error) possible in these commonly used products.

The next chapter of this thesis examines the methodology used to determine the GIS product accuracies derived from the data layers used in this case study. After a discussion of inherent error assessment and sampling methodology, the creation and recording of the thematic maps is discussed. By comparing the results of fieldwork to map data, accuracy evaluation matrices were created. Operational and composite error assessment are presented at the end of the chapter.

CHAPTER III

METHODOLOGY

A concern shared by all who create and/or use GIS products is that of the accuracy of the data sources and accuracy of the resulting products. Using a questionnaire dealing with data quality , Mead (1982) examined responses from 17 managers of state level geographic information systems. While 10 of these managers admitted to definite data quality problems in their systems, and all 17 believed it is always or sometimes possible to measure data quality, only 9 of the 17 said that statements regarding data quality and/or appropriate uses of the data are ever included on their GIS products. As pointed out by Mead (1979) and Vitek, Walsh, and Gregory (1984), such accuracy statements need to be included on any base product or final GIS product in order that the usefulness of such items may be assessed. Accuracy refers to both position and content of information, and concerns errors inherent in base products as well as operationally induced errors in final GIS products. Accuracy can be stated in either statistical or absolute terms but never based on subjective opinions (Mead, 1982).

This chapter examines the procedures used within this thesis to quantify the degree of accuracy possible with any GIS product produced within this case study. Every thematic map overlay used in a GIS will have some combination of both locational and non-locational errors. Because the highest accuracy possible in any GIS product can only be, at best, equal to that of the least accurate map layer (Newcomer and Szajgin, 1984), the accuracy assessment of individual map layers to be used in a GIS is the logical place to begin final product assessment. Sections discussing the steps used to determine base product accuracies (inherent error assessment) as well as operational error assessment procedures are described below.

Sampling Procedure

Hay (1979) suggests sample sizes of 50 or larger while VanGenderen, Lock, and Vass (1977 and 1978) show that a sample size as low as 30 can be used to derive 90 percent interpretation accuracy at the 0.05 significance level, which meets the U.S. Geological Survey criterion of the accuracy for land use interpretation. VanGenderen, Lock, and Vass (1978) further state that the likelihood of making incorrect interpretations on insufficient samples must be realized and accuracy levels adjusted accordingly. The sample size used in this study was placed at 35 points in order to allow for additional points beyond the minimum of 30 recommended by VanGenderen, Lock, and Vass (1977 and 1978). Additional sample points were not needed because intra-class data comparisons, requiring 30+ sample points per class, were not to be addressed by this study.

Congalton (1984) found that simple random sampling performs quite well when testing the spatial complexity of landcover maps despite the strong spatial autocorrelation often found in different images. When identifying agriculture and forest areas, in fact, simple random sampling is preferred over stratified random sampling, which seems to work best for identifying range differences (Congalton, 1984). A simple random sample was, therefore, used for choosing the position of the 35 sample points. Although some sources (VanGenderen et al., 1978; Hay, 1979; Ginevan, 1979) point out that stratified random sampling techniques are often accepted as the most appropriate in resource studies, such a sampling procedure was not used because each sample point was to be used for simultaneously recording landcover, terrain, and soils data. Therefore, even if the sample points were stratified for one map source, they would then be non-stratified on all other maps because the placement of each point is at the same geographical position on each map. In addition, the study area was small enough in size as to eliminate the need to stratify the study area by access routes. The number of samples per slope angle, slope aspect, landcover, or soil type categories was not particularly important to the purposes of this study because critical evaluation of the mapping accuracy of each class was not the goal. The objective was to compare map generated information to field collected information. Detailed map category or class accuracy assessment is beyond the scope of this study. This analysis seeks only to assess a percent accuracy of all 35 points sampled within each thematic map. Data were recorded only for cells that were geographically referenced to the same sample points in order to evaluate the probabilities of a cell in one location being classified accurately at the same point within each layer in the GIS. Such a process can be used to determine combined error accuracies in accordance with the concepts discussed in Newcomer and Szajgin (1984).

With sample sizes as low as 1 to 5 per class, which occur in this study, interpretation accuracies would not be reliable for the testing of significant relationships if one was examining the classification accuracies of individual classes (VanGenderen, 1978). Therefore, such intra-class interpretations are not addressed by this study. An attempt only was made to assess the overall accuracy of each thematic map layer, thereby allowing all 35 sample points to be used only for making statements concerning the overall accuracy of a map product. The 35 sample points, then, allow a 90 percent or better interpretation accuracy significance at the 0.05 level as pointed out by VanGenderen (1978).

In two cases, slope aspect and soil type, only 34 sample points were used. Because one of the sample points was located in water, that point was not used. The point was retained, however, because it was still relevant to landcover and slope angle classification. In the case of soil type, one sample point identified in the field by flags was removed from its location by cattle.

Random sample points were determined by using a random numbers table to generate 35 pairs of X and Y coordinates. This method allows for the unbiased selection of all sample points. These points were plotted onto a mylar grid which was divided into a 50 by 50 matrix. The outside boundaries of the grid corresponded to the 4 square mile (10.4 sq. km) study area on the 1:24,000 scale topographic and thematic maps. The minimum sample point spacing on the 50 x 50 cell matrix was 1/25th of a mile or 211.2 feet (64.4 m). This level of spacing was used to prevent points from clustering any closer together than 200 feet (60 m), thereby providing maximum dispersion of sample points across the entire 2 x 2 mile (10.4 sq.km) study area. This grid was overlaid on the study area on a 1:24,000 scale topographic map and each sample point was plotted onto this map. Each sample point was given a reference number of 1 to 35 (corresponding to the 35 sample points) in the order that it was plotted. All numbers making reference to sample points within this thesis relate to this numbering scheme.

To aid in the locating of sample points in the field, the 35 points identified on the 7.5 minute series topographic map were transferred to a set of August 1982 black and white aerial photographs at a scale of 1:3,600. These were the most recent large scale aerial photographs of the study area available at the time that this study was done. All sample points were transfered through the use of a Bausch and Lomb zoom transfer scope.

Creation and Recording of Thematic Map Data

Four different thematic maps of the study area were created specifically for this thesis so that each map could be compared to field data and thereby assessed for accuracy. The separate maps are: (1) landcover (Figure 4); (2) slope angle (Figure 5 and 6); (3) slope aspect (Figure 7 and 8); and (4) soils (Figure 9). All maps were produced at 1:24,000 scale because of the common use of this scale in GIS and related work. The use of the same scale throughout each layer (including those data sets which had to be converted to 1:24,000) allowed for the manual overlay of all thematic map layers so that all sample points could be geographically referenced to the same location on each map. The errors found to be inherent in these commonly used GIS data layers could then be used

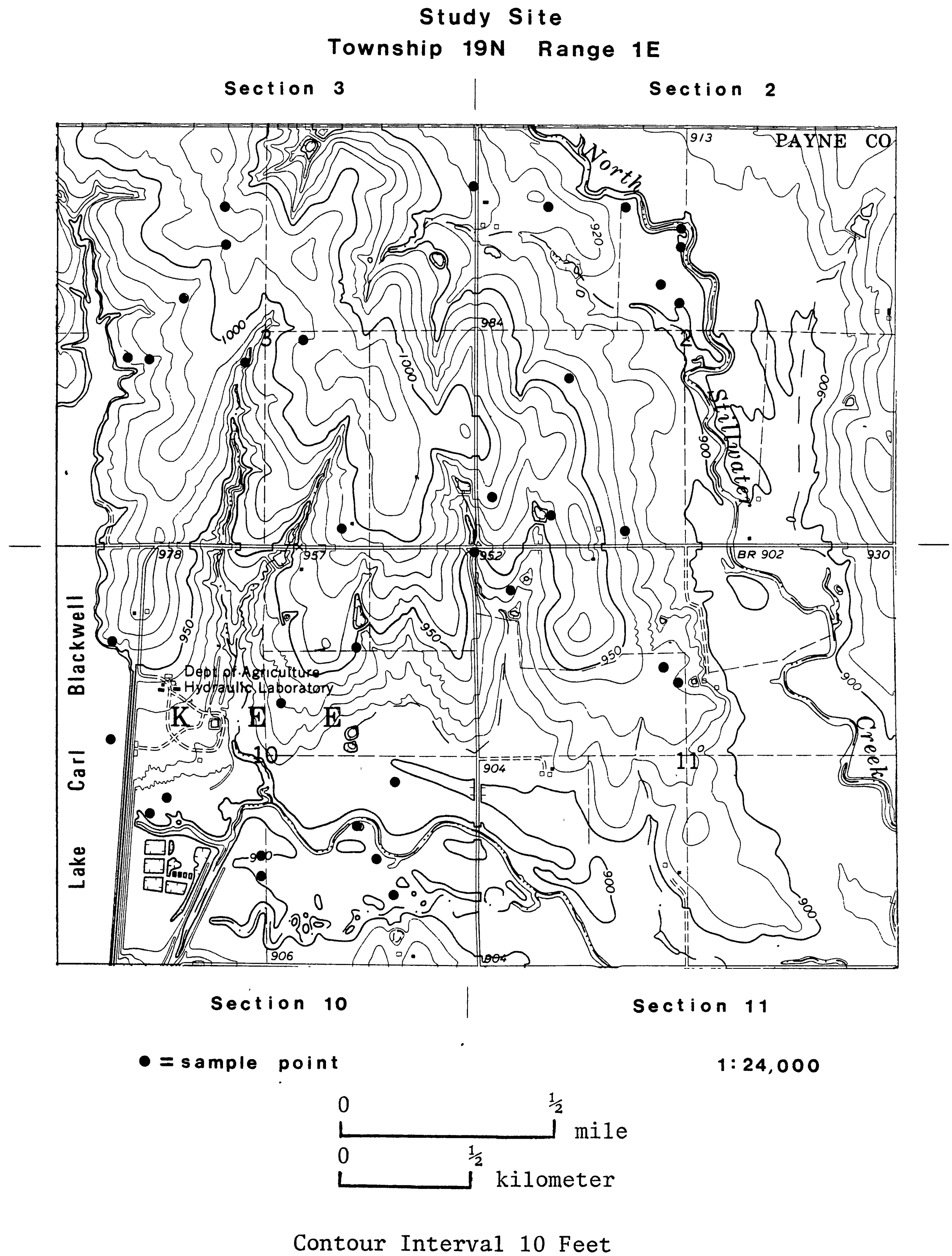


Figure 3. Topographic Map of Study Area

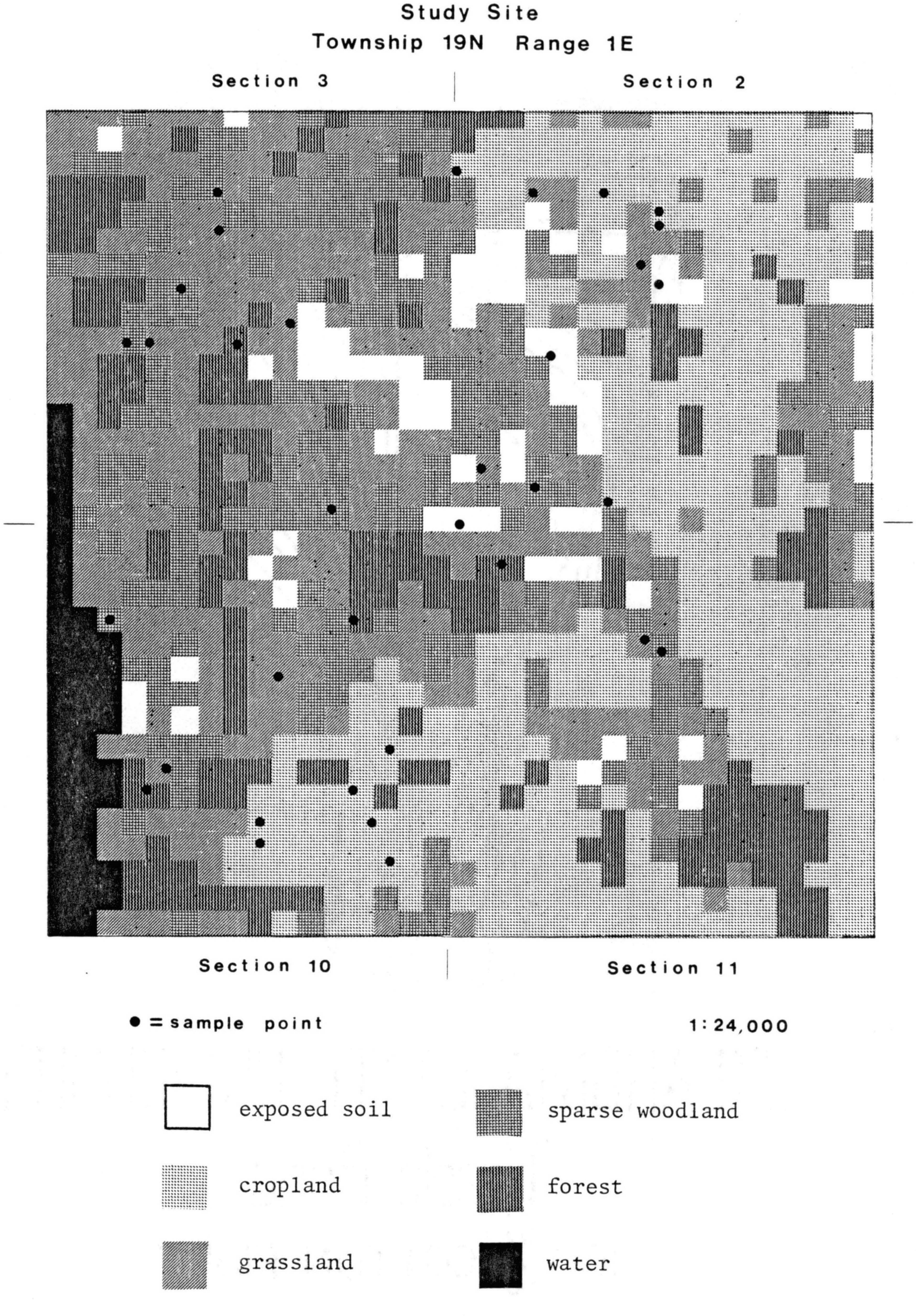


Figure 4. 2.5 Acre (100 m^2) Landcover Map

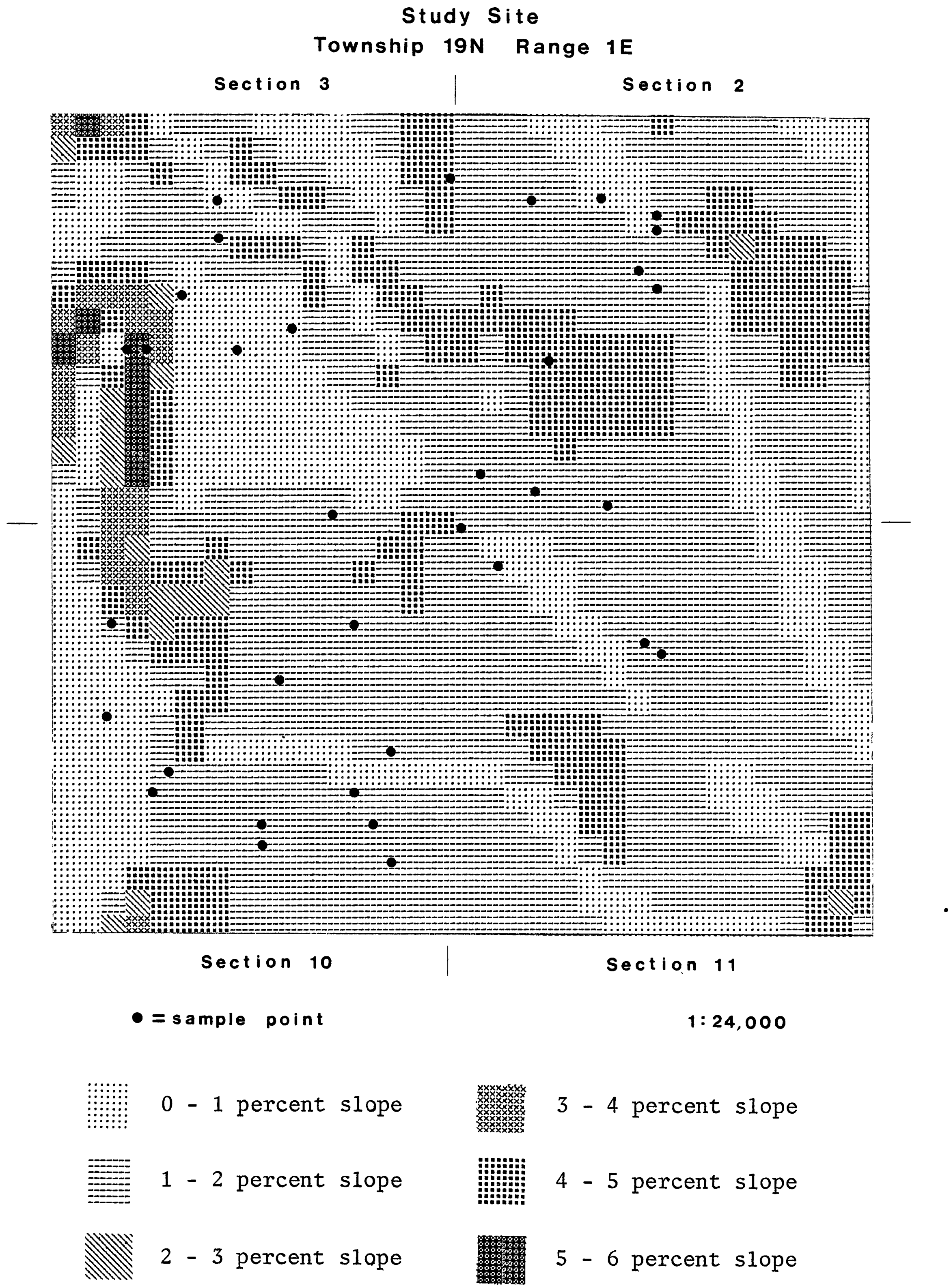


Figure 5. 2.5 Acre (100 m^2) Slope Angle Map

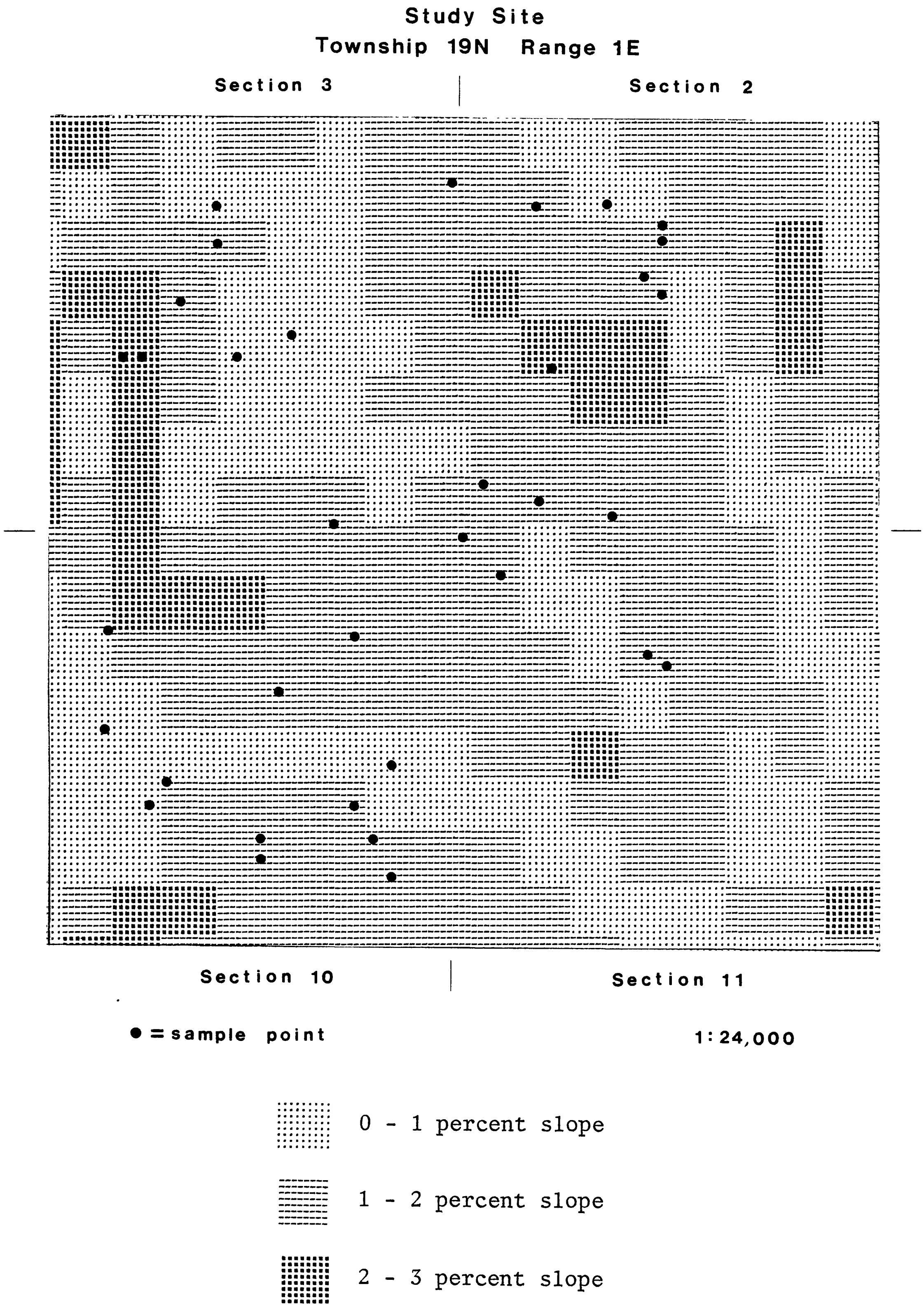


Figure 6. 10.0 Acre (200 m^2) Slope Angle Map

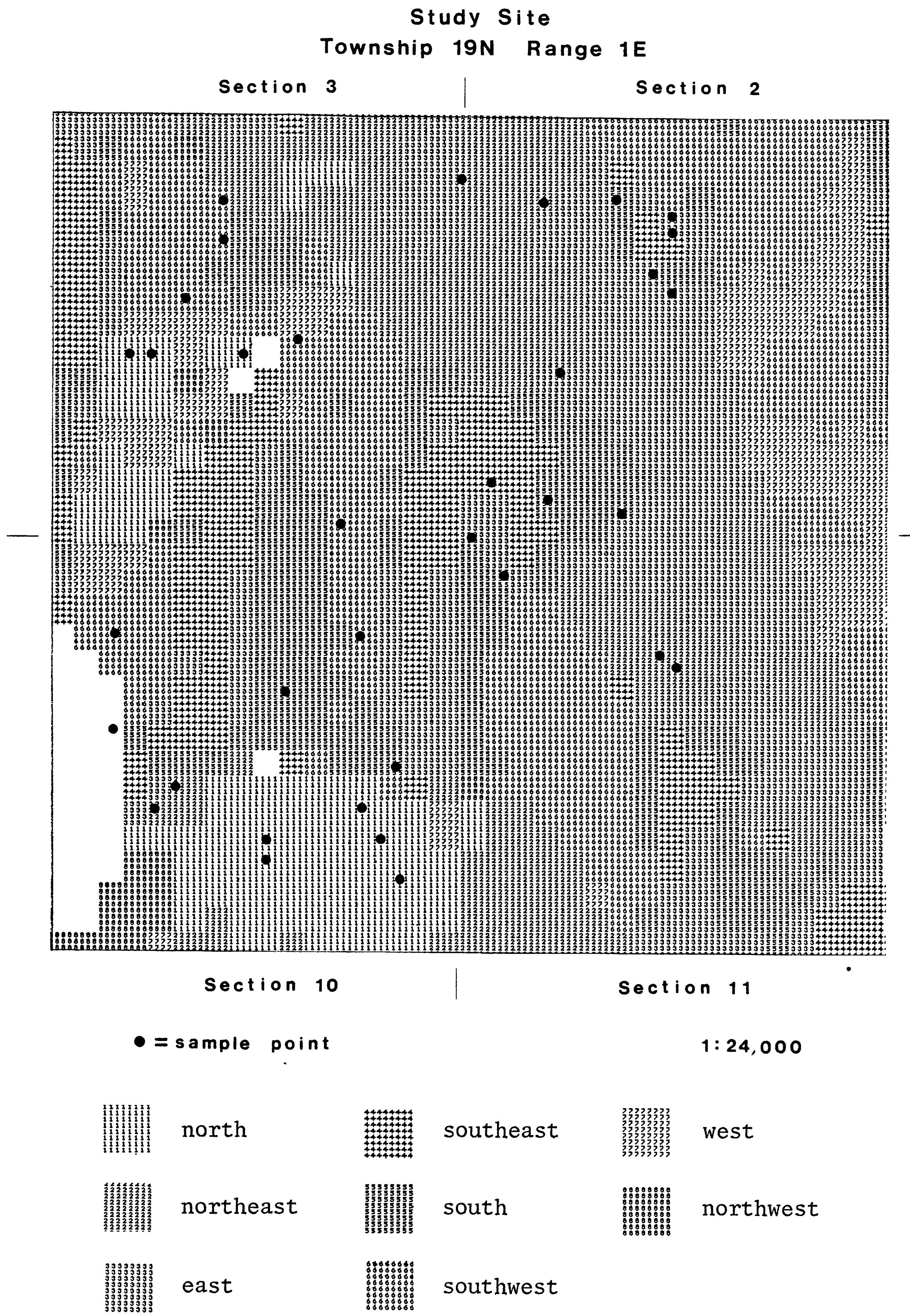


Figure 7. 2.5 Acre (100 m^2) Slope Aspect Map

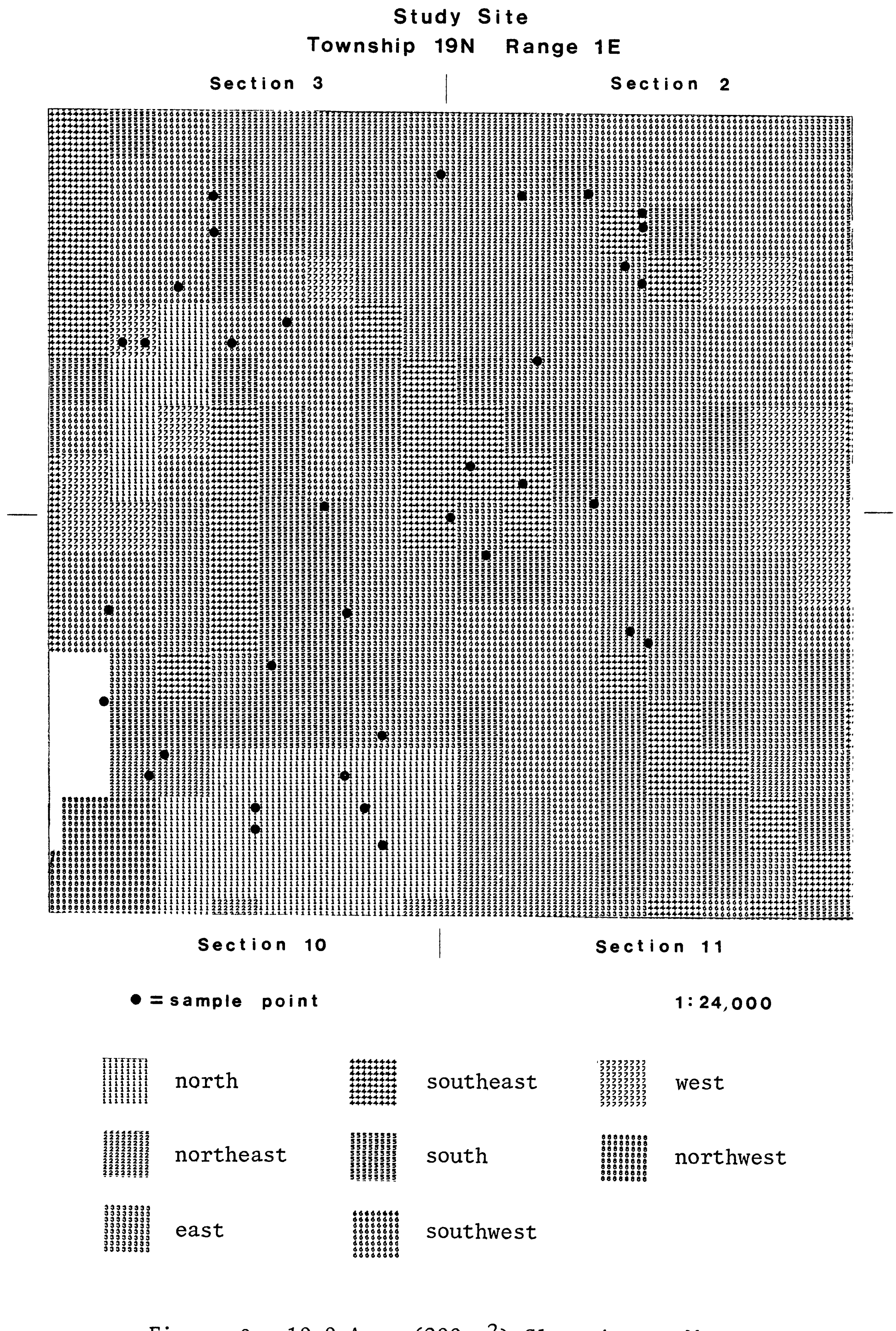
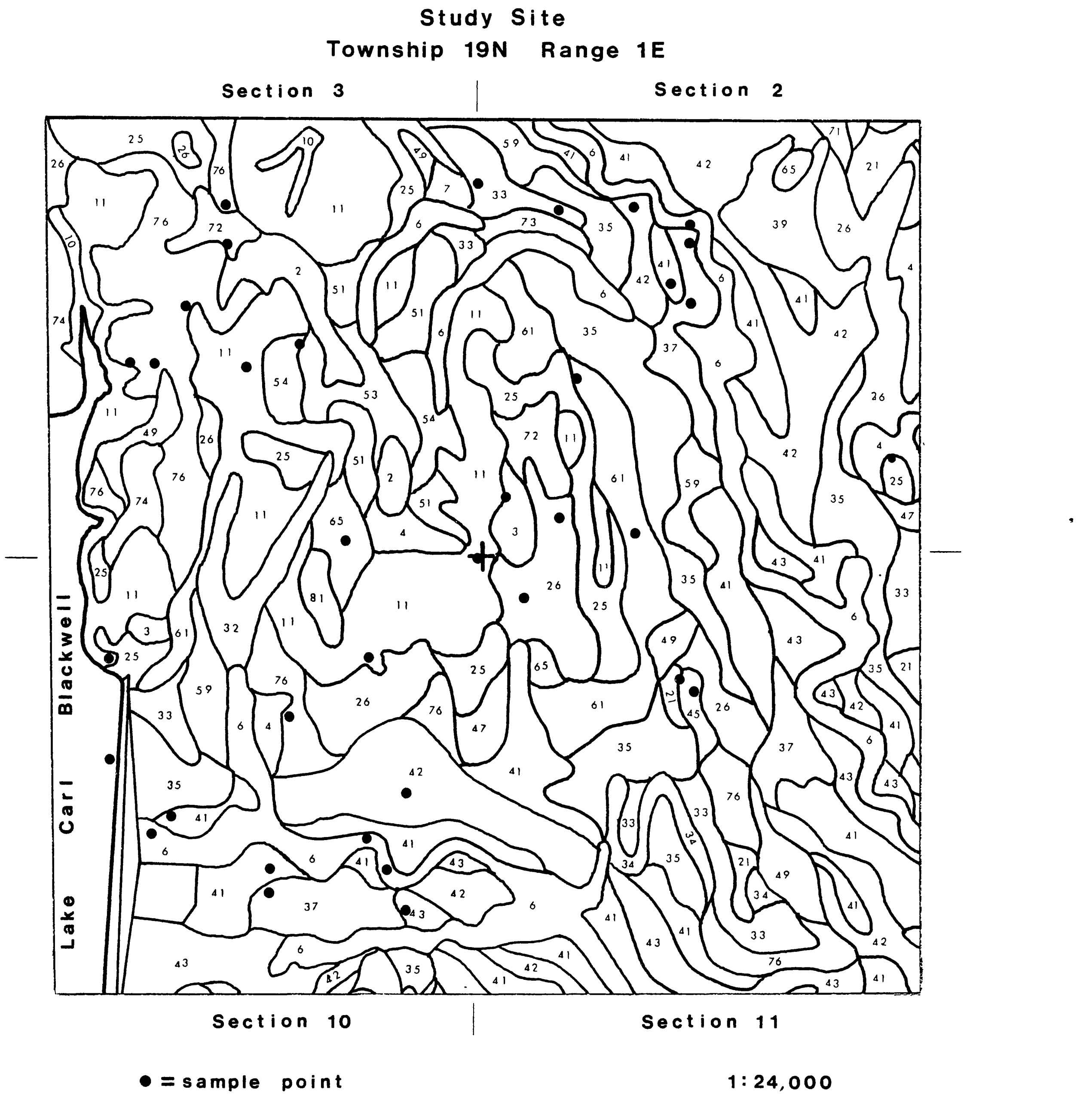


Figure 8. 10.0 Acre (200 m^2) Slope Aspect Map



* Refer to Appendix C for numeric soil unit legend.

Figure 9. Soils Map of Study Area

to address the operational and composite error problems to be discussed later in this chapter. Because most computer based GIS's are designed to process grid data (Vitek, Walsh, and Gregory, 1984), all maps were prepared using a grid cell format rather than polygon delineations. A comparison of the operating costs and final accuracies of products using cell or polygon systems of data capture shows that, while polygon data capture produces a slightly higher spatial accuracy, a cellular grid system is much more computer compatible, is eight to ten times less expensive, and is much faster to use than polygons (Wehde, 1982).

Because a cellular system forces the selection of a grid cell size to determine the resolution of any resulting map, all maps in this study were produced at both a 2.5 acre (100 square meter) and a 10.0 acre (200 square meter) resolution in order to assess the role that these different cell sizes have upon inherent error. The problem of cell size selection has been previously addressed by Wehde (1982), Henderson (1980), Gersmehl and Gersmehl (1982), and Muller (1977) and results from this thesis will be compared to their findings. For the soil type map, additional distinction was made by producing two maps using the cell midpoint method of data capture (one for each of the two grid resolution sizes) and another map for each cell size area using cell dominant data capture to assess the comparative accuracy of these two methods of data capture.

Landcover Maps

The most common application for computer interpretation and mapping using Landsat data is landcover or vegetation cover mapping (Wilson and Thomson, 1982). An 8 April 1981 Landsat-4 multispectral scanner digital data set of the study area, obtained from the Center for Applications of Remote Sensing (CARS) at Oklahoma State University, was classified. A more recent tape would have been desirable but, of all Landsat tapes on file at CARS, the April 1981 tape was the nearest in date to the August 1982 aerial photographs used for fieldwork. More recent air photos of the study area were not available from stock at the large scale desired. The landcover types found within the study area are fairly static and, therefore, would not be expected to change much during the time elapsed between the MSS tape and aerial photograph dates, or the MSS tape and fieldwork dates. The 1:3,600 scale aerial photography was used as an ancillary data source to help determine delineations of landcover types. One map was produced for a 2.5 acre (100 sq.m) and a 10.0 acre (200 sq.m) grid cell resolution. The landcover classes identified are the Level I landcover types of the U.S. Geological Survey land use classification system (Anderson et al, 1976). The identified landcover classes are: (1) water; (2) cropland; (3) exposed soil; (4) range or grassland; (5) sparse woodland (less than 50 percent crown closure); and (6) forest (50 percent or greater crown closure). These classes were used because of their simplicity in classification both on the landcover map and in the field. Furthermore, the poor resolution of the 10.0 acre (200 sq.m) grid landcover map prevents the practical use of any landcover types more detailed than the six classes used. After completing the two landcover maps on the Comtal image processor at CARS, they were printed at a scale of 1:24,000 on an electrostatic printer (Figure 4). A 10.0 acre (200 sq.m) resolution landcover map is not included as a figure in this thesis because that map source could be printed only as digital matrix data because of system problems. After overlaying and geographically referencing the 1:24,000 scale mylar overlay of the sample points (previously described), landcover types were recorded for each sample point on each of the two landcover maps (Appendix A). If any sample point fell on a boundary line between any two grid cells, the value of the cell to the north and east (in that order) was taken.

Terrain Maps

A 1:250,000 scale digital terrain tape covering the study area was obtained and processed at CARS. Although a 1:24,000 scale terrain tape would provide greater accuracy, a digital terrain tape at this scale has not yet been made available by USGS for the study area. The 1:24,000 scale terrain products have presently been created only for areas of mountainous terrain or for special study sites. Studies may even utilize terrain maps of greater accuracy than that possible with the 1:24,000 scale digital elevation models. More accurate terrain products may be produced by manual digitizing or automated scanning of elevation values at closely spaced intervals from large scale topographic maps or stereopair aerial photographs. Because manual digitization is up to five times less accurate than automated digitizing (Chrisman, 1982a), creates higher running costs, is inherently slow, and produces many errors contributed to the presence of a human operator (Marble and Peuquet, 1983), automated digitized products are frequently used. Only in certain aspects of existing map work, however, can the high speed and economic advantages of automated digitizing be realized (Marble and Peuquet, 1983). Because of the operational difficulties involved in automated scanning, any organization should give serious consideration to the question of using its own automated scanner. High quality results can only be obtained in a dedicated situation where the scanner is kept fully operational, so smaller users should use other facilities and service bureaus. The question must be asked "Is the digitizing of contour sheets required or are the data not better obtained from automatically digitized digital elevation models (DEM's)" (Marble and Peuquet, 1983). Further rationale for the choice of commercially available digital products over individually produced automated terrain models is provoded by Marble, Lauzon, and McGranaghan (1984) who report that a current lack of software support systems for spatial data often leaves a researcher with only crude tools to undertake the task of spatial data handling. This, in turn, often leads the researcher to the personal decision not to use automated digitizing for the handling of spatial data (Marble, Lauzon, and McGranaghan, 1984). Environmental planning and management professions, therefore, rely extensively on data collected and produced by specialized agencies or other groups who have previously studied a site (Sinton, 1978). Higher accuracy terrain products would obviously contribute to greater accuracy within any product utilizing these terrain maps, yet the widely available 1:250,000 digital terrain tapes, being both faster and easier to obtain, are often used within spatial data systems. Examples of other studies utilizing the 1:250,000 scale digital terrain tapes in a GIS are Klock, Gum, and Jordan (1986), Root et al (1986), Hart, Wherry, and Bain (1985), Walsh, Stadler, and Gregory (1984), and Walsh and Stadler (1983). Because the 1:250,000 scale terrain data used for this thesis are so commonly used is GIS systems, and this thesis attempts only to analyze a typical GIS procedure, a 1:250,000 scale USGS digital terrain tape is used within this study.

The terrain tape was used to produce separate maps for both slope angle and slope aspect; each of these was in turn produced at grid cell resolutions of 2.5 acres (100 sq.m) and 10.0 acres (200 sq.m). Various levels of percent slope, patterned after the classes used with a 1:250,000 scale digital terrain tape in Walsh and Stadler (1983), were represented on the two slope angle maps and classed as: 0-1, 1-2, 2-3, 3-4, 4-5, and 5-6 percent slope. Slopes steeper than six percent were not recognized by the terrain tape within the study area because the landscape is relatively flat and thus very few contour lines were found at such a small map scale. The classes used within the slope aspect map, also modeled after Walsh and Stadler (1983), represented the eight primary points of the compass and are: (1) north--337.5-22.5 degrees; (2) northeast--22.5-67.5 degrees; (3) east--67.5-112.5 degrees; (4) southeast--112.5-157.5 degrees; (5) south--157.5-202.5 degrees; (6) southwest--202.5-247.5 degrees; (7) west--247.5-292.5 degrees; and (8) northwest--292.5-337.5 degrees.

In applications where the scale and resolution of the intended use is known, it makes sense to rescale digital maps (reducing the number of points used to represent a digitized line) to this known scale (VanHorn, 1985). Maps are often produced at the 1:24,000 scale so that they may be used at the same scale as aerial resource photos or USGS topographic maps (Klock, Gum, and Jordan, 1986), so all four maps used in this study were subsequently output on hardcopy at 1:24,000 scale using an electrostatic printer (Figure 5,6,7, and 8). The 1:24,000 scale sample point plotting grid was then overlaid and geographically referenced to these terrain maps in order to record the slope angle and the slope aspect indicated by the maps for each of the 35 sample points (Appendix A). If any sample point fell on a division between cells, the class of the slope aspect or slope angle to the north and east was recorded.

Soils Maps

Data input into a GIS are often originally produced as irregular polygon areas which need to be converted to a grid format that is more usable by a computer (Vitek, Walsh, and Gregory, 1984). A computer generated soil type map is one such type of polygonal data source commonly used within geographic information systems. Detailed soils maps are manually coded for computerization by visually selecting the dominant soil within a unit cell or grid and then entering the information into a computer (Nichols, 1975). Because this process introduces error into resultant maps by reducing the detail and number of soil mapping units and by changing the delineations of soil mapping units from the original map (Nichols, 1975), such a process is necessary for input into computers and subsequent comparisons to other layers of GIS data. A simulated computerized soil type map was created for this study by visually selecting and manually coding soils information from a soils map without actually inputing these data into a computer.

A 1:20,000 scale Soil Conservation Service county soil survey of the study area was photographically reduced to the 1:24,000 scale used by all products throughout this study (Figure 9). The sample point grid was overlaid and geographically referenced to the soil type data. The soil type at each of the 35 points was subsequently recorded (according to the SCS soil series mapping units of the area) for the following situations: (1) soil type at each sample point for the 2.5 acre (100 sq.m) cell midpoint grid value and the 10.0 acre (200 sq.m) cell midpoint grid value; and (2) 2.5 acre (100 sq.m) cell dominant area grid value and 10.0 acre (200 sq.m) cell dominant area grid value (Appendix A). The soil types recorded at each point serve as the basis for assessing the accuracy of the original SCS soils map. The difference in cell size and method of soil type designation within each cell provides information regarding data capture techniques and the resulting impact on map accuracy. Any sample points falling on a line between cells were assigned to the cell to the north and east. This direct overlay method of converting original polygon data to a grid cell format does make some unrealistic assumptions about the distribution of soils (or other data) over a surface, but it has the advantage of being relatively simple to perform and is therefore used in many GIS (Markham and Townshend, 1981).

Field Data Collection

In order to determine the extent of errors inherent in all thematic maps layers created for this study, the landcover, slope angle and aspect, and soil type recorded at each sample point on the maps needed to be checked against the actual conditions present at each of the 35 control points. Because most of the study area was privately owned, the first step involved going through tax records to get the names and phone numbers of all landowners in the area for the purpose of requesting access to their properties. Of the 13 different landowners/land managers on whose properties all original 35 sample points were located, two refused access so additional sites were identified for use. Before going into the field a listing of angles and distances from permanent surface features to each sample point was prepared from the large scale (1:3,600) aerial photographs of the study area using protractor and ruler measurements. Permanent features used were corners of known buildings, fencerow junctions, road intersections, and water towers. These reference points then served as locations from which to measure distances and angles to sample points in the field. The fieldwork was completed in five days and was broken down into two phases.

Phase one consisted of physically locating all 35 sample points in the field and recording landcover, slope angle, and slope aspect at each site (Appendix B). Precise location of each point in the field was accomplished with the aid of a survey transit for measuring azimuths and a 100 foot (30.5 m) tape measure to determine distances. Initial points of reference used to locate all points were those determined from the large scale aerial photos as described above. After setting up the transit at each starting reference point, both azimuth and distance measures were performed to precisely locate each sample point. At least two back-azimuth measures were made to known permanent surface features from each sample point to confirm the location of that point. Once found, landcover, slope angle, and slope aspect values were recorded at each point. Landcover was recorded according to the six class scheme previously described. Landcover, slope angle, and slope aspect measures were recorded as relative point values (as opposed to larger areas) by observing and/or measuring the landcover, slope angle and slope aspect within the immediate six to ten foot (two to three meter) area surrounding each sample point. Slope angle and aspect were determined by a Brunton compass placed on a six foot (two meter) section of 2x4 inch (five by ten cm) board laid across the ground in the direction of the maximum slope of each point. A Brunton compass was used because both slope angle and slope aspect may be measured with the same instrument. Back-azimuth measures shot from each point usually indicated less than six foot (two meter) horizontal placement error of each point. Before leaving each site, a numbered flag was placed in the ground for the purpose of relocating each sample point during the soils portion of the fieldwork.

Phase two consisted of recording the soil type at each of the 35 field sample points (Appendix B). The classification of soils was performed in the field by Mr. Jim Henley, Perry Office, Oklahoma Soil Conservation Service. Mr. Henley personally mapped the soils in the study area during a recent SCS soils mapping revision of Payne County, Oklahoma. After the flag at each sample point was located, the point was cored by either a hand auger or a hydraulic soil probe and the soil core was subsequently analyzed according to the SCS soil series mapping legend for the area (Appendix C). The soil type found to be at each point was recorded regardless of whether it was a major mapping unit or only an inclusion soil within a mapping unit. Although soil inclusions are acknowledged and described in text form within an SCS Soil Survey, the exact delineations are not shown on soils maps and are therefore not digitized into any GIS data layer.

Matrices for Accuracy Evaluation

In order to assess the errors inherent in each of the maps produced for this study, a series of 11 matrices for accuracy evaluation were produced by comparing field recorded data to map recorded data (Table I, II, III, and IV). Such evaluation matrices are commonly employed for assessing the classification accuracies of landcover maps by comparing field sampled values to the Landsat derived classification to create an accuracy measure for each landcover class and to assess overall map accuracy (Walsh, Vitek, and Gregory, 1982). When using matrices for accuracy evaluation, the reference or field data are compared to the classified map data and the major diagonal across the matrix indicates the agreement between field and map data. Overall accuracy for a classified map is calculated by dividing the sum of the entries that form the major diagonal (the number of correct classifications) by the number of samples taken (Story and Congolton, 1986). The same format used for the landcover evaluation matrices presented in Walsh, Vitek, and Gregory (1982) and Story and Congolton (1986) were used as a model for all evaluation matrices in this study in order to compare landcover, slope angle and aspect, and soils map information to those values found during fieldwork. The classification accuracy of individual classes of landcover, terrain, and soils is not important to the

TABLE I

LANDCOVER ACCURACY EVALUATION MATRICES

2.5 acre (100 m^2) Resolution

FIELDCHECKED LANDCOVER

LANDCOVER MAP	Exposed Soil	Cropland	Range	Sparse Woodland	Forest	Water
Exposed Soil	1	2				
Cropland		5		2	3	
Range		3	5	1		
Sparse Woodland			4	4		
Forest					4	
Water						1

20 of 35 sample points correctly assigned

Overall Accuracy: 57.1%

* Underlined digits on all accuracy matrices indicate the number of correctly classified sample points per data class.

TABLE I (Continued)

10.0 acre (200 m^2) Resolution

LANDCOVER MAP (rows) / FIELDCHECKED LANDCOVER (columns)

	Exposed Soil	Cropland	Range	Sparse Woodland	Forest	Water
Exposed Soil	0	1	1			
Cropland		4			3	
Range	1	5	3	2		
Sparse Woodland			3	4	1	
Forest			2		3	
Water				1		1

15 of 35 sample points correctly assigned

Overall Accuracy: 42.9%

TABLE II

SLOPE ANGLE ACCURACY EVALUATION MATRICES

2.5 acre (100 m^2) Resolution

DIGITAL TERRAIN TAPE MAP	FIELDCHECKED SLOPE ANGLE %					
	0-1	1-2	2-3	3-4	4-5	5-6>
0-1	1		2	1		2
1-2	2	4		2	3	14
2-3			0			
3-4				0		
4-5		1			1	
5-6						2

8 of 35 sample points correctly assigned

Overall Accuracy: 22.9%

TABLE II (Continued)

10.0 acre (200 m^2) Resolution

DIGITAL TERRAIN TAPE MAP	FIELDCHECKED SLOPE ANGLE %					
	0-1	1-2	2-3	3-4	4-5	5-6>
0-1	2		2	1		2
1-2	1	4		2	4	14
2-3		1	0			2
3-4				0		
4-5					0	
5-6						0

6 of 35 sample points correctly assigned

Overall Accuracy: 17.1%

TABLE III

SLOPE ASPECT ACCURACY EVALUATION MATRICES

2.5 acre (100 m^2) Resolution

DIGITAL TERRAIN TAPE MAP	FIELDCHECKED SLOPE ASPECT							
	N	NE	E	SE	S	SW	W	NW
N	1			2	3	1	1	
NE		1		2	1		1	1
E	1		1		2	2		1
SE				3				
S			1		2		1	
SW			2			1		
W							3	
NW								0

12 of 34 sample points correctly assigned

Overall Accuracy: 35.3%

TABLE III (Continued)

10.0 acre (200 m^2) Resolution

DIGITAL TERRAIN TAPE MAP	FIELDCHECKED SLOPE ASPECT							
	N	NE	E	SE	S	SW	W	NW
N	1			2	1		1	
NE		1		2			1	1
E			1		1	1		1
SE				3	1	1	1	
S	1		3		4		1	
SW						2		
W					1		2	
NW								0

14 of 34 sample points correctly assigned

Overall Accuracy: 41.2%

TABLE IV

SOILS ACCURACY EVALUATION MATRICES

Point Recorded Soils Data from SCS SOILS MAP[1]

FIELDCHECKED SOILS[2]

SCS SOILS MAP	2	3	4	6	11	25	26	33	35	37	39	41	42	43	45	51	54	59	61	65	72	76	Water
2	–																						
3		–																			1		
4			–																				
6				1								1	2										
11					4																		
25						–																	
26							1												*1				
33								2															
35									–														
37										1		1	1	1									
39											–												
41												2	1										
42												1	1										
43														–									
45		1													1								
51																–							
54																	1						
59																		–					
61																			2				
65						1														–			
72																					1		
76					2																	3	
Water																							1

<u>Without inclusions</u>: 20 of 34 sample points correctly assigned

Overall Accuracy: 58.8%

<u>With inclusions</u>: 33 of 34 sample points correctly assigned

Overall Accuracy: 97.1%

[1]See APPENDIX C for SCS Soil numbers legend.

[2]*Means that a field checked soil is not an accepted inclusion within the soils mapping unit.

TABLE IV (Continued)

Cell Midpoint - 2.5 acre (100 m^2) Resolution

FIELDCHECKED SOILS

SCS SOILS MAP	2	3	4	6	11	25	26	33	35	37	39	41	42	43	45	51	54	59	61	65	72	76	Water
2	–																1						
3		–			1																1		
4			–																				
6				–								1		1									
11					2																	1	
25					1	–													1				
26		1					–								1								
33								1															
35								1	–			1							2				
37										–		1	2										
39											–												
41				1								2											
42													3										
43														–									
45															–								
51																–							
54																	–						
59																		–					
61																			–				
65						1														–			
72																					1		
76					2																	2	
Water							1																1

12 of 34 sample points correctly assigned

Overall Accuracy: 35.3%

TABLE IV (Continued)

Cell Midpoint - 10.0 acre (200 m^2) Resolution

FIELDCHECKED SOILS

SCS SOILS MAP	2	3	4	6	11	25	26	33	35	37	39	41	42	43	45	51	54	59	61	65	72	76	Water
2	–																						
3		–			2																		
4			–			1																	
6				–				1				2	2										
11					2														1		1	1	
25						–																	
26							1								1						1	1	
33								–															
35		1							–			1							2				1
37										–													
39											–			1									
41				1								1											
42												1	3										
43														–									
45															–								
51																–	1						
54																	–						
59								1										–					
61																			–				
65																				–			
72																					–		
76					2																	1	
Water							1																1

8 of 34 sample points correctly assigned

Overall Accuracy: 23.5%

TABLE IV (Continued)

Cell Dominant - 2.5 acre (100 m^2) Resolution

FIELDCHECKED SOILS

SCS SOILS MAP	2	3	4	6	11	25	26	33	35	37	39	41	42	43	45	51	54	59	61	65	72	76	Water
2	—																1						
3		—			1																1		
4			—																				
6				—										1									
11					2																	1	
25					1	—													1				
26		1					1								1								
33								1															
35								1	—			1							2				
37										—		1	2										
39											—												
41				1								3											
42													3										
43														—									
45															—								
51																—	1						
54																	—						
59																		—					
61																			—				
65						1														—			
72																					1		
76					2																	2	
Water																							1

14 of 34 sample points correctly assigned

Overall Accuracy: 41.2%

TABLE IV (Continued)

Cell Dominant - 10.0 acre (200 m^2) Resolution

FIELDCHECKED SOILS

SCS SOILS MAP	2	3	4	6	11	25	26	33	35	37	39	41	42	43	45	51	54	59	61	65	72	76	Water
2	–																1						
3		–			1																		
4			–			1																	
6				–								2	1										
11					3																1	1	
25						–																	
26		1					1								1							1	
33								–															
35				1				1	–			1							2				1
37										–			1										
39											–												
41												1											
42												1	3	1									
43														–									
45															–								
51																–							
54																	–						
59								1										–					
61																			1				
65																				–			
72																					1		
76					2																	1	
Water																							–

11 of 34 sample points correctly assigned

Overall Accuracy: 32.4%

focus of this study. The goal was to determine only an overall classification accuracy statement for each thematic map as a whole.

The evaluation matrices were designed to assess the accuracy of each of the 11 thematic map products. Two landcover evaluation matrices were produced for the 2.5 acre (100 sq.m) and the 10.0 acre (200 sq.m) resolution landcover maps; two more for the 2.5 acre (100 sq.m) and 10.0 acre (200 sq.m) resolution slope angle maps; and another two for the same two resolution versions of the slope aspect maps. Five evaluation matrices of soil type were created to examine: (1) soils recorded at a point on the SCS soils map; (2) 2.5 acre (100 sq.m) cell midpoint map; (3) 10.0 acre (200 sq.m) cell midpoint map; (4) 2.5 acre (100 sq.m) cell dominant map; and (5) 10.0 acre (200 sq.m) cell dominant map.

Initial evaluation of all matrices took the form of examining the percent error thus shown to be inherent in each of the 11 map products. Visual comparisons were also made between the 2.5 acre (100 sq.m) and 10.0 acre (200 sq.m) map evaluations, and the effects that cell midpoint versus cell dominant area methods of data capture had on mapping accuracies. The results of all of these accuracy evaluations are discussed in detail in chapter IV.

Operational and Composite Error Assessment

In previous sections of this chapter, inherent error in GIS was assessed by checking the accuracy of each thematic map against field data. Operational error will be assessed by combining two or more thematic map layers, in various combinations, to test for the statistical probability of composite error which results from such data layering. Operational errors may be categorized into two major types: (1) positional errors and (2) identification errors (Newcomer and Szajgin, 1984). Positional errors, which occur from inaccuracies in the horizontal placement of cell boundaries, are addressed by Henderson (1980) and Wehde (1982). Although every product will have some combination of both operational error types, alignment errors are assumed to be negligible within this study since the data gathered are best suited for looking at identification errors. Identification errors occur from the mislabeling of areas of the various categories on thematic maps (Newcomer and Szajgin, 1984). Additional sources of operational errors possible in a GIS can occur because of human error in digitizing information, GIS algorithm inaccuracies, and human bias when categorizing and delineating data on computerized maps.

Statistical analyses of operational error follows the concepts outlined by Newcomer and Szajgin (1984). They point out that as the number of layers in a GIS increases, the number of possible error combinations increases considerably. In order for a correct assignment to result in any particular grid cell on a GIS product, every vertically aligned cell in each data layer used in the GIS must also be correctly assigned (Figure 10). Any other combination of alignment that includes even one of the aligned and geographically referenced grid cells in the data stack will result in an incorrect assignment in that cell (Figure 11). Therefore, the highest accuracy possible in any GIS product can only be equal to the accuracy of the least accurate individual map layer and the lowest possible accuracy is equal to the sum of all incorrectly assigned cells in each data layer used. This worst case could occur when all mislabeled cells are found at different locations throughout all data layers.

Statistical results of various combinations of two and three layer GIS products were computed to determine the amount of identification-type operational error which could be present in such products. This operational error assessment was done by setting up a table showing the 35 sample points in one column and landcover, slope angle, slope aspect, and soil type in another column. Both 2.5 acre (100 sq.m) and 10.0 acre (200 sq.m) versions of each of these thematic maps were analyzed. Table V and VI presents the accuracy assessment for a 2.5 acre (100 sq.m) and 10.0 acre (200 sq.m) cell, respectively. The sample

COMBINED ERRORS in GIS

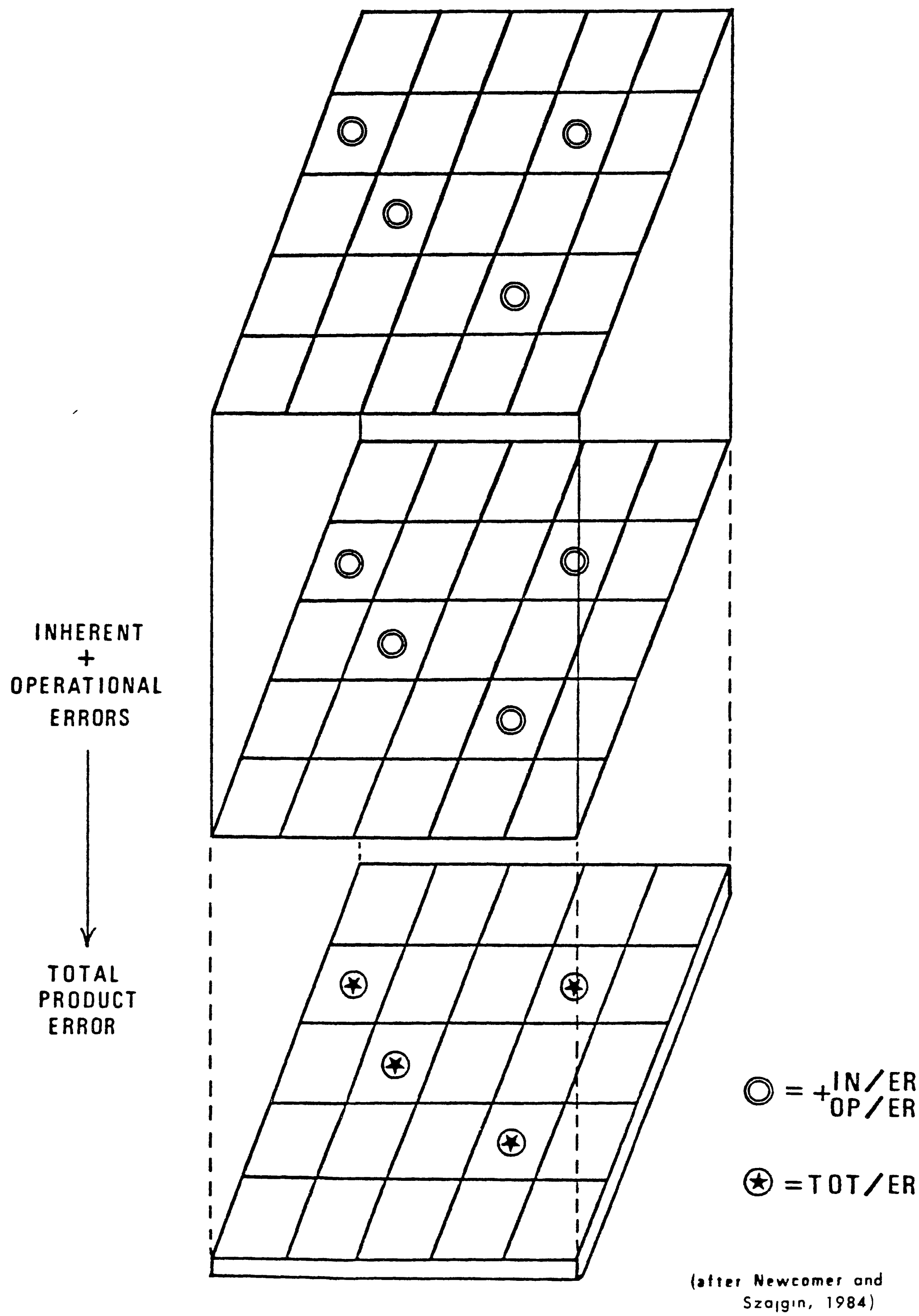


Figure 10. Best Case Combined Error

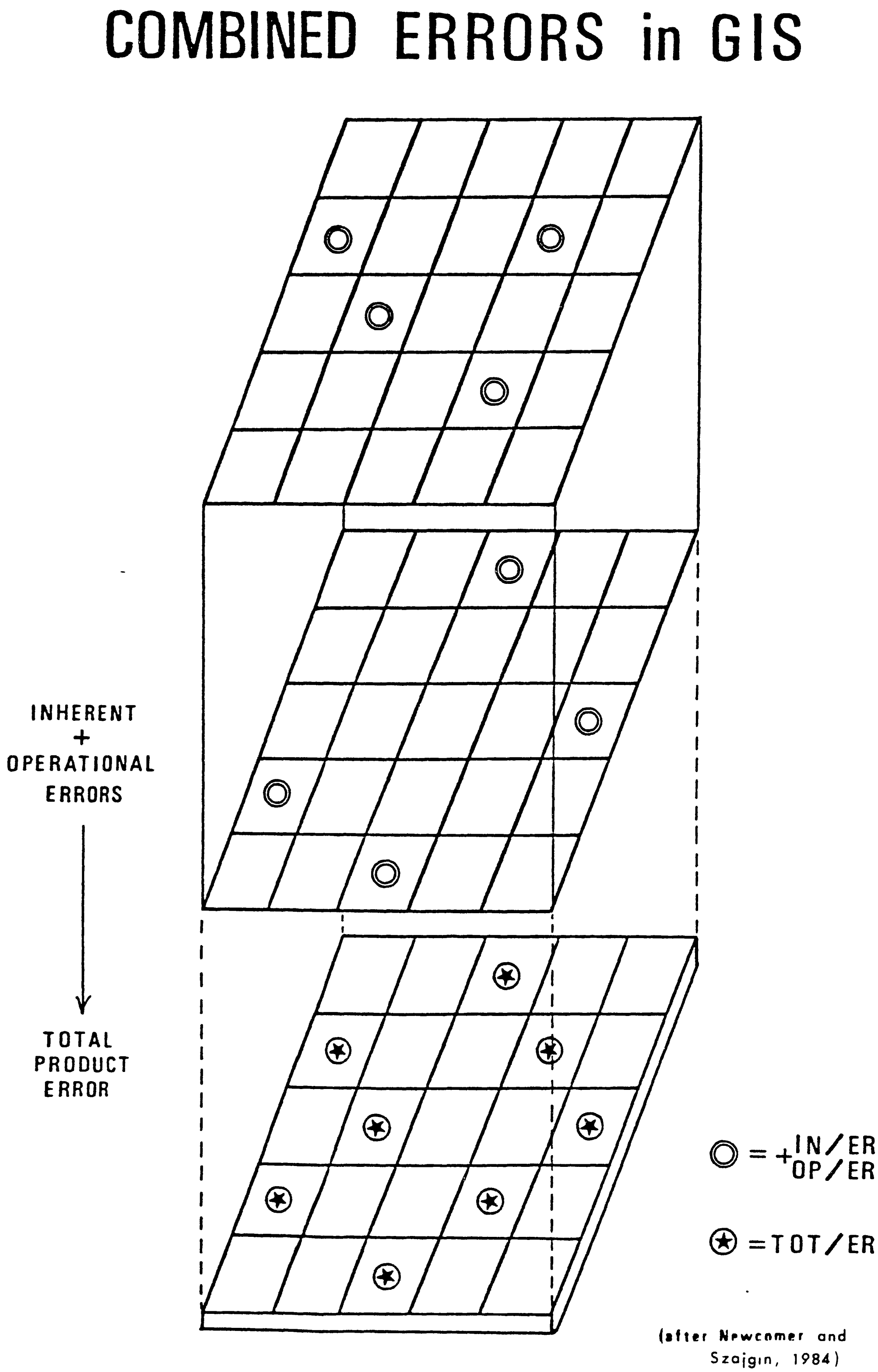


Figure 11. Worst Case Combined Error

points on each map were labeled on these tables as either an "X" (incorrectly labeled point) or an "O" (correctly labeled point). Theoretical product accuracies were determined by counting how many sample points had correctly labeled cells throughout each layer of the data stack analyzed. Even one incorrectly labeled cell in a stack resulted in an error for that point on the GIS product. Any sample points not classified, for reasons previously mentioned, were counted as an "O" by default (Table V and VI). The theoretical upper and lower accuracy limits were also computed for each GIS product, according to the concepts and formulae given in Newcomer and Szajgin (1984). The lower limit is calculated as $1 - \sum_{i=1}^{n} Pr(Ei)$, where n = the number of layers used in the GIS, i = data layer 1,2,3...n, and Pr(E) = the probability of error in a data layer. The upper limit is calculated as the maximum Pr(Ei), or the percent accuracy of the least accurate map in the GIS layer. All of these results are discussed in detail in chapter IV which follows.

TABLE V

ANALYSIS OF 2.5 ACRE (100 m^2) CELL RESOLUTION

(X = incorrect cell assignment per sample point)

(O = correct cell assignment per sample point)

SAMPLE POINT #	LANDCOVER	SLOPE ANGLE	SLOPE ASPECT	SOILS
1	X	O	O	O
2	X	X	X	X
3	O	O	X	X
4	O	X	X	O
5	X	X	X	O
6	O	X	O	O
7	O	X	X	X
8	X	O	O	X
9	O	X	O	X
10	X	O	X	X
11	X	X	O	X
12	O	O	X	X
13	X	X	X	O
14	O	X	X	O
15	X	X	X	X
16	O	X	O	X
17	X	X	X	X
18	O	X	O	X

TABLE V (Continued)

(X = incorrect cell assignment per sample point)

(O = correct cell assignment per sample point)

SAMPLE POINT #	LANDCOVER	SLOPE ANGLE	SLOPE ASPECT	SOILS
19	X	X	O	X
20	O	X	X	X
21	O	X	O	O
22	O	X	O	X
23	O	X	X	O
24	X	X	X	X
25	X	X	X	X
26	O	X	O	O
27	O	O	not used	O
28	O	O	X	O
29	O	O	X	X
30	O	X	X	X
31	X	X	X	O
32	X	X	O	O
33	O	X	X	O
34	X	X	X	X
35	O	X	X	not used

TABLE VI

ANALYSIS OF 10.0 ACRE (200 m^2) CELL RESOLUTION

(X = incorrect cell assignment per sample point)

(O = correct cell assignment per sample point)

SAMPLE POINT #	LANDCOVER	SLOPE ANGLE	SLOPE ASPECT	SOILS
1	X	X	O	X
2	X	X	X	X
3	O	O	X	X
4	O	X	X	X
5	X	X	X	O
6	O	X	X	X
7	X	X	O	X
8	X	O	O	X
9	O	X	O	X
10	X	O	X	X
11	X	X	O	X
12	O	X	O	X
13	O	X	X	O
14	O	X	O	O
15	X	X	X	X
16	X	X	O	X
17	X	X	X	X
18	O	X	X	X

TABLE VI (Continued)

(X = incorrect cell assignment per sample point)

(O = correct cell assignment per sample point)

SAMPLE POINT #	LANDCOVER	SLOPE ANGLE	SLOPE ASPECT	SOILS
19	X	X	O	X
20	X	X	O	X
21	O	X	O	O
22	X	X	O	X
23	X	O	X	O
24	X	X	X	X
25	O	X	X	X
26	X	X	O	X
27	O	O	not used	X
28	O	X	X	O
29	X	O	X	X
30	O	X	X	X
31	X	X	X	O
32	X	X	O	X
33	O	X	X	O
34	X	X	X	X
35	O	X	X	not used

CHAPTER IV

DATA ANALYSIS

In order to achieve wider acceptance among users of thematic maps and GIS products created as a result of combining such maps, an interpreter must be able to specify product accuracy. The goal of this chapter is to present and explain such derived accuracy measures. Initially, inherent error will be assessed since all subsequent error quantification is a direct result of the accuracies (or inaccuracies) of thematic base maps. Analysis of the accumulation of operational and composite errors will follow.

Quantification of Inherent Error

Assessment of inherent error is concerned foremost with deriving statistical accuracies of landcover, slope angle, slope aspect, and soils maps used for this study. Secondly, because it is evident that the particular method of cell mapping and the cell size employed will have an effect on the accuracy and validity of resulting thematic maps, comparisons between all of the 2.5 acre (100 sq.m) and 10.0 acre (200 sq.m) products and between cell dominant area and cell midpoint generated maps are made.

Landcover maps proved to be the most accurate layer of data analyzed, although the best landcover map only had a final accuracy of 57.1 percent, with 20 of the 35 sample points correctly assigned to a landcover class. It is expected that the accuracy of this map could be improved by reassigning classes with consideration to the discrepancies found in the field, but only the initial classification of landcover will be analyzed at this time. A 14.2 percent drop in classification accuracy resulted from the comparison between the 2.5 acre (100 sq.m) resolution landcover map (57.1 percent accuracy) and the 10.0 acre (200 sq.m) resolution landcover map (42.9 percent accuracy). This represented the largest resolution-dependent drop found as a result of such comparisons (Table VII). This large accuracy drop is likely because of the fine resolution of the original Landsat data being highly generalized through aggregation of landcover values at the coarse 10.0 acre (200 sq.m) grid resolution.

Slope angle maps presented the lowest accuracies of all four thematic map types analyzed, most likely because of the high probability of finding small undulations in slope when recording the field data at a point. The best digital slope angle map produced was the 2.5 acre (100 sq.m) version and had only 8 of the 35 sample points, or 22.9 percent, accurately classified. The 10.0 acre (200 sq.m) resolution version dropped 5.8 percent in accuracy

TABLE VII

ANALYSIS OF INHERENT ERROR

Evaluation Type	Resolution	Accuracy	# Sample Points Correct
1. Landcover Map Accuracies:	2.5 acre (100 m^2)	= 57.1%	(20/35)
	10.0 acre (200 m^2)	= 42.9%	(15/35)
2. Slope Angle (Terrain Tape) Accuracies:			
	2.5 acre (100 m^2)	= 22.9%	(8/35)
	10.0 acre (200 m^2)	= 17.1%	(6/35)
3. Slope Aspect (Terrain Tape) Accuracies:			
	2.5 acre (100 m^2)	= 35.3%	(12/34)
	10.0 acre (200 m^2)	= 41.2%	(14/34)
4. Soils Map Accuracies:			
At-a-point	without inclusions	= 58.8%	(20/34)
	with inclusions	= 97.1%	(33/34)
Cell midpoint	2.5 acre (100 m^2)	= 35.3%	(12/34)
	10.0 acre (200 m^2)	= 23.5%	(8/34)
Cell dominant	2.5 acre (100 m^2)	= 41.2%	(14/34)
	10.0 acre (200 m^2)	= 32.4%	(11/34)
5. 2.5 acre (100 m^2) vs. 10.0 acre (200 m^2) Accuracies: (Totals of landcover, terrain, soils)			
	2.5 acre (100 m^2) average	= 38.4%	(66/172)
	10.0 acre (200 m^2) average	= 31.4%	(54/172)
6. Cell Dominant vs. Cell Midpoint Accuracies: (Totals of both 2.5 (100 m^2) and 10.0 (200 m^2) acre soils data)			
	Cell dominant average	= 36.8%	(25/68)
	Cell midpoint average	- 29.4%	(20/68)

with only 17.1 percent of the 35 sample slope angle points accurately identified from the 1:250,000 scale digital terrain tape (Table VII). Slope aspect maps were considerably more accurate than slope angle, because the aspect at the micro-scale still often followed the trend of the large-scale slope aspect. The most accurate slope aspect map was the 10.0 acre (200 sq.m) resolution landcover map which attained a final accuracy of only 35.3% (Table VII). This slope aspect map represented the only case in this study where a decrease in the resolution of data aggregation resulted in an increase in the accuracy of that product. In all other cases (landcover, slope angle, and soils) the decrease in resolution resulted in an expected accuracy drop. This increase in accuracy was perhaps because the 10.0 acre (200 sq.m) resolution is closer to the resolution of the original 1:250,000 scale digital terrain tape.

The soil maps allow examination of resolution dependent and data aggregation methods. The first map tested against field data was the actual 1:20,000 scale SCS soils map used to produce the grid soil type maps. Soils observed in the field at each point were compared to the soils indicated at each point on the SCS soils map. The accuracy of this product was determined to be 58.8 percent, with 20 of the 34 sample points correctly assigned (Table VII). Most of this inaccuracy is likely the result of "striking" soil inclusions within regular soil series mapping units during field sampling. Soils maps would obviously prove more accurate if such inclusions could be allowed for within the soils map or any digitized products derived from such a map. Table VII, therefore, also assesses the accuracy of the original SCS soils map (97.1 percent) by allowing for soil inclusions to determine the accuracy of this product as commonly used in cases where the mapping or digitizing of such inclusions is of little significance. At the scale the map was originally produced (1:20,000), however, SCS has determined that any inclusions four acres (1.61 hectares) or smaller in area cannot be accurately delineated and therefore are only mentioned in text form and not mapped (SCS, 1984)(Table VIII). Because such inclusions are not mapped, they cannot be digitized for use with a GIS or any spatial analysis study, and are therefore excluded from consideration within this study, except as an initial assessment of the original soil type map as made available by SCS.

The best digital soils map produced was the 2.5 acre (100 sq.m) resolution map aggregated by the cell dominant method with 14 of the 34 points correctly labeled for an accuracy of 41.2 percent. The 10.0 acre (200 sq.m) resolution version of these cell dominant products dropped to a 32.4 percent accuracy. The maps aggregated by cell midpoint had accuracies of 35.3 percent and 23.5 percent for the 2.5 acre (100 sq.m) and 10.0 acre (200 sq.m) versions, respectively (Table VII).

TABLE VIII

SOILS MAPPING SCALES AND MINIMUM DELINEATION SIZE

Map Scale		Inches per mile	Minimum size delineation[1] (acres)	(hectares)
1:500		126.7	0.0025	0.001
1:2,000		31.7	0.040	0.016
1:5,000		12.7	0.25	0.10
1:7,920		8.00	0.62	0.25
1:10,000		6.34	1.00	0.40
1:12,000		5.28	1.43	0.58
1:15,840		4.00	2.5	1.01
1:20,000		3.17	4.0	1.62
1:24,000	(7½′)	2.64	5.7	2.31
1:31,680		2.00	10.0	4.05
1:62,500	(15′)	1.01	39	15.8
1:63,360		1.00	40	16.2
1:100,000		0.63	100	40.5
1:125,000		0.51	156	63.2
1:250,000		0.25	623	252
1:300,000		0.21	897	363
1:500,000		0.127	2,500	1,012
1:750,000		0.084	5,600	2,267
1:1,000,000		0.063	10,000	4,048
1:5,000,000		0.013	249,000	100,809
1:7,500,000		0.008	560,000	226,720

[1] The "minimum size delineation" is taken as a 1/16 sq.in. (0.4 sq.cm.) area. Cartographically, this size is about the smallest area in which a symbol can be printed readily. Smaller areas can be delineated, and the symbol lined in from outside, but such very small delineations drastically reduce map legibility.

(From SCS Soil Survey Manual, 1984)

Final comparisons between the effects of cell size on overall accuracies and between cell dominant and cell midpoint aggregated products show definite decreases in accuracy with reduced resolution and a drop in accuracy when aggregating map data by grid cell midpoint. Out of a combined 172 sample points tested on all four thematic map types, 66 of these, or 38.4 percent, were correctly classified on all 2.5 acre (100 sq.m) resolution maps. In contrast, only 54 of the 172 points were correct on all 10.0 acre (200 sq.m) version maps, representing an average drop in accuracy of 7.0 percent attributable to a decrease in resolution (Table VII). Comparisons of cell dominant and cell midpoint accuracies, performed only with the soils data, resulted in a correct assignment for 25 of the 68 combined points (36.8 percent) used by both resolution versions of the cell dominant method maps. Only 20 of these 68 points (29.4 percent) used in both of the cell midpoint maps were accurately labeled. This represents an average drop in accuracy of 7.4 percent when using the cell midpoint method of data capture (Table VII).

Quantification of Operational and Composite Error

To assess the amount of operational error in a GIS product, combined errors present in such multilayer data sets are first determined. Because the highest accuracy possible using multilayer data stacks is equal to the accuracy of the least accurate individual map layer (Newcomer and Szajgin, 1984), operational error is calculated as the difference between total combined error actually existing in a GIS product and the least accurate layer used within that product. Operational error, then, is that error responsible for the drop in accuracy found between inherent error (which goes into the GIS) and composite error (which is found in products coming out of the GIS).

When every erroneous point in each of the map layers being used occurs in the same location throughout the data stack, the theoretical maximum accuracy possible in a multilayer data set will occur. Conversely, the lowest accuracy that can result is obtained when the errors in each map layer occur at unique point locations throughout the stack. In probabilistic terms, these errors are ocurring at mutually exclusive, disjoint locations (Newcomer and Szajgin, 1984). Knowledge of such accuracy values allows both upper and lower limits to be determined concerning the accuracy of a composite map. By counting the number of sample points which were correctly assigned throughout each layer in a data stack, actual composite map accuracies of such products were assessed. Using the concepts and formulae described by Newcomer and Szajgin (1984), theoretical upper and lower limits (based on probability theory) were calculated for each GIS product as well. Table IX shows these actual composite errors found within each hypothetical GIS product created by combining various thematic maps into two and three-layer data sets. This table also shows both the theoretical upper and lower limits to product accuracy, calculated as described above.

The most accurate GIS product possible, using the thematic maps produced for this study, is a two-layer map created by combining the 2.5 acre (100 sq.m) resolution landcover map and the 2.5 acre (100 sq.m) cell dominant soils maps. The actual error that would be present in such a product was determined to be 71.4 percent. In other words, only 10 of the 35 sample points on the 2.5 acre (100 sq.m) resolution maps had both landcover and soils correctly labeled for a combined accuracy of only 28.6 percent. Upper and lower limits to accuracy were calculated as 41.2 percent and 2.9 percent, respectively (Table IX). The most accurate three-layer product possible is a combination of 2.5 acre (100 sq.m) resolution versions of landcover, slope aspect, and cell dominant area soils maps. The actual accuracy of such a map would be only 11.4 percent (four of 35 points correctly classified throughout all three layers), while upper and lower limits are respectively calculated as 35.3 percent and 0.0 percent, respectively (Table IX). Other combinations of two and three-layer data sets were also used and the results of all such multilayer product

TABLE IX

ANALYSIS OF COMBINED ERROR

	2-Layer Products	Theoretical Lower Limit[1]	Actual Composite Map Accuracy[2]	Theoretical Upper Limit[3]
1.	2.5 acre (100 m^2) Landcover and Cell Dominant Soils	2.9% accuracy (1/35)	28.6% accuracy (10/35)	41.2% accuracy (14/34)
2.	10.0 acre (200 m^2) Landcover and Cell Dominant Soils	0.0% (0/35)	17.1% (6/35)	32.4% (11/34)
3.	2.5 acre (100 m^2) Cell Dominant Soils and Slope Aspect	0.0% (0/35)	17.1% (6/35)	35.3% (12/34)
4.	10.0 acre (200 m^2) Cell Dominant Soils and Slope Aspect	0.0% (0/35)	11.4% (4/35)	32.4% (11/34)
5.	2.5 acre (100 m^2) Landcover and Slope Aspect	0.0% (0/35)	22.9% (8/35)	35.3% (12/34)

TABLE IX (Continued)

3-Layer Products	Theoretical Lower Limit[1]	Actual Composite Map Accuracy[2]	Theoretical Upper Limit[3]
1. 2.5 acre ($100\ m^2$) Landcover, Slope Aspect, and Cell Dominant Soils	0.0% accuracy (0/35)	11.4% accuracy (4/35)	35.3% accuracy (12/34)
2. 10.0 acre ($200\ m^2$) Landcover, Slope Aspect, and Cell Dominant Soils	0.0% (0/35)	5.7% (2/35)	32.4% (11/34)

[1]Calculated as $[1 - \sum_{i=1}^{n} Pr\ (Ei)]$, where n = the number of layers used in the GIS, i = data layer 1, 2, 3...n, and Pr(E) = the probability of error in a data layer (after Newcomer and Szajgin, 1984).

[2]Calculated by manually aligning georeferenced cells according to TABLES I and II.

[3]Calculated as the maximum (Pr (Ei)), or the percent accuracy of the least accurate map in the GIS layer (after Newcomer and Szajgin, 1984).

accuracies are shown in Table IX. Only one combination of a four-layer set resulted in a product having an accuracy above 0.0 percent. Such a product was created by combining all four 2.5 acre (100 sq.m) resolution maps and had only 1 of the 35 sample points correctly labeled throughout the four thematic maps, resulting in a very low accuracy of 2.9 percent. Upper and lower limits were 22.9 percent and 0.0 percent, respectively. All other four-layer combinations resulted in true error rates of 100 percent.

By subtracting the results of true combined error accuracies from those of the least accurate layer found within that data stack, operational error could be addressed. This is done by subtracting composite map accuracy from the theoretical upper limit, since this upper limit is actually the accuracy of the least accurate map used in the process (Newcomer and Szajgin, 1984). In other words, operational error is that error which reduces the accuracy of a GIS map from its theoretical best to the level of error actually possessed. Operational error for the representative combinations of data layers previously shown in Table IX are calculated and presented in Table X. A two-layer combination of 2.5 acre (100 sq.m) resolution landcover and cell dominant soils yields a 12.6 percent increase in error between the least accurate map in this data set (the soils map) and the true composite error derived; thus, an operational error of 12.6 percent degraded the map combination to the level of total error indicated. A three-layer data set comprised of the 2.5 acre (100 sq.m) resolution landcover, slope aspect, and cell dominant soils maps resulted in an operationally induced error of 23.9 percent (Table X).

It can be shown, therefore, that an 11.3 percent drop in accuracy resulted between the two-layer data set and a three-layer data set using similar map products. A similar comparison of 10.0 acre (200 sq.m) resolution versions shows a two-layer set of landcover and soils (operational error of 15.3 percent) dropping in accuracy by 11.4 percent just by adding an additional layer to the process (landcover, slope aspect, and soils: operational error of 26.7 percent)(Table X). These results clearly support the contention made by Newcomer and Szajgin (1984) that error increases rapidly as the number of layers used in composite maps increases. All data presented in this chapter are summarized and presented along with conclusions to the study in the next chapter.

TABLE X

ANALYSIS OF OPERATIONAL ERROR

*Operational error = Theoretical Upper Limit Accuracy - Actual Composite Accuracy

2-Layer Products	Operational Error
1. 2.5 acre (100 m^2) Landcover and Cell Dominant Soils	12.6%
2. 10.0 acre (200 m^2) Landcover and Cell Dominant Soils	15.3%
3. 2.5 acre (100 m^2) Cell Dominant Soils and Slope Aspect	18.2%
4. 10.0 acre (200 m^2) Cell Dominant Soils and Slope Aspect	21.0%
5. 2.5 acre (100 m^2) Landcover and Slope Aspect	12.4%

3-Layer Products	Operational Error
1. 2.5 acre (100 m^2) Landcover, Slope Aspect, and Cell Dominant Soils	23.9%
2. 10.0 acre (200 m^2) Landcover, Slope Aspect, and Cell Dominant Soils	26.7%

CHAPTER V

RESULTS AND CONCLUSIONS

A summary of the results determined through data analysis is as follows:

(1) Landsat landcover maps (57.1 percent accuracy in the best case) proved to have less inherent error than any of the grid-cell terrain or soils maps. Both the terrain and the soils maps had best case accuracies of 41.2 percent.

(2) Slope angle maps presented the lowest accuracies of any of the four thematic map products used, with a best case accuracy of 22.9 percent.

(3) In all but one case, smaller grid cell sizes (2.5 acre or 100 sq.meter resolution) resulted in higher accuracies when compared to larger grid cell sizes (10.0 acre or 200 sq.meter resolution).

(4) An average increase in accuracy of 7.0 percent resulted from the use of smaller grid cells.

(5) Cell dominant method of data capture proved to be more accurate than cell midpoint, with an average improvement in accuracy of 7.4 percent.

(6) The best two-layer GIS product used within this study is made up of 2.5 acre (100 sq.m) resolution landcover and cell dominant soils maps and resulted in a combined accuracy of 28.6 percent.

(7) The best three-layer GIS product was obtained using 2.5 acre (100 sq.m) versions of landcover, slope aspect, and cell dominant soils maps, resulting in a combined accuracy of only 11.4 percent.

(8) Creation of any product using all four data layers at any resolution produced resulting GIS products at or near 100 percent error. This inaccuracy was largely caused by the limiting accuracy of the slope angle map.

(9) Combined error increased dramatically as the number of stacked layers increased. Any products produced from a combination of more than two layers had very poor accuracies.

(10) Operational error followed the same trends as combined error: increasing as cell size increased, as less accurate products were utilized, and as the number of layers in the GIS increased.

The major focus of this research assesses the actual error found to be inherent in the base products used in this study. Based upon the accuracies derived for each GIS data overlay, accuracy statements for a resultant GIS product are provided along with minimum and maximum accuracy levels possible for each combination of data layers. Because this study utilizes only four variables within a limited geographical area having gently rolling topography and uniform landcover, studies covering larger areas of more complex terrain and utilizing more variables will only increase the accumulation of error within the final product. While the Landsat landcover map produced by this study was the most accurate of the four types of map products tested, improvements in accuracy still could have been made by using a more recent MSS tape of the area rather than one that was more than a year older than the aerial photographs used and more than three years older than the fieldwork performed.

Because fieldwork recorded data at specific points, one of the biggest problems with misclassification resulted from edge effects in which, for example, a point recorded in the field as range is adjacent to a large area of forest, the whole vicinity being classified as forest on the relatively low resolution landcover map. The fundamental problem is that the grid cells must be included or excluded in their entirety as there is no spatial information at the sub-grid cell level (Crapper, 1984). The map space occupied by the sample point dot represents approximately ten feet (three meters) in this study. This size presents little problem, however, because landcover was largely determined by examining the immediate six to ten foot (two to three meter) area surrounding the field sample point, and terrain data were recorded along a six foot (two meter) segment. The original resolution of the Landsat data and the 1:24,000 scale these data were mapped at would also contribute to error. Although maps as large as 1:25,000 scale can be made from Landsat MSS data, the resolution of the original data (1.1 acres or 79 m) usually restricts its use to large area surveys at smaller scales (Wilson and Thomson, 1982).

The low levels of accuracy found to be inherent in products created from the digital terrain data is likely because of the series of generalizations of original surface data that results from the creation of these slope angle and slope aspect maps. Relatively small scale (1:250,000) maps, already grossly generalized and thus possessing considerable inherent error, are further generalized by sampling and digitizing points at only 200 foot (60 m) intervals off of the original topographic map in order to produce the digital terrain tapes. These digital data are then further removed from reality during the aggregation of the terrain tape data into grid cells during the creation of the slope angle and slope aspect maps.

Soil type map accuracies were below only those of landcover maps. The original SCS soils map of the study area had an accuracy of 58.8 percent and, if allowance is made for soil inclusions, this accuracy increases to 97.1

percent. Although inclusions are acknowledged within the text of the soil survey, such enclaves of similar and dissimilar soils cannot be mapped because of their small areas relative to the scale at which the map was created (Table VIII). Because these inclusions are not mapped, they cannot be digitized and are therefore excluded from any GIS decision making process. The 1:20,000 scale at which the original SCS map was created allows the omission of soil inclusions as large as 4+ acres (1.61 hectares) in size constituting up to 20 percent of the total area of any soil association. Soil complexes within such a mapping scale may have similar sized inclusions making up as much as 25 percent of a total mapping unit without including such soils on the map (Henley, 1985). Sample points "landing" within these non-mappable inclusions were therefore fairly common and contributed greatly to the amount of error detected within these soil products. The accuracy of soil type maps used by a GIS will depend on the area under study because the accuracy of such products varies significantly with the complexity of the areas' soils, the experience of the mapper, and the availability and use of ancillary data sources such as color infra-red and stereopair airphotos (Henley, 1985). The results obtained by the data in this thesis agree with the results of Nichols (1975) that the accuracy of computerized soil type maps drops with increases in the cell size used to digitize the map.

The selection of a smaller cell size was shown to have a definite influence on improving map accuracies, which agrees with similar findings by Monmonier (1983), Hay (1979), VanGenderen et al(1978), and Wehde (1982). Improvement of the slope aspect map accuracy when using a larger cell size likewise agrees with Henderson (1980). He found that the smallest cell size did not consistently generate the most precise data in every case. The rapid accumulation of both combined and operational errors found to exist because of increases in the numbers of map layers in a GIS agrees with like results discussed in the error accumulation study by Newcomer and Szajgin (1984).

All maps were assessed only as whole units because the accuracy of individual categories or classes within a thematic map was not important to the purposes of this study. With the sample scheme utilized, any statements made concerning individual class accuracies would not be valid because too few samples "per class" were taken and, therefore, no opportunity exists to accurately test for intra-class differences. Results of this study might be altered slightly by a different cell placement or by rotation of cells to a different position, as shown by Henderson (1980), but such factors as positional or alignment errors must be assumed to be negligible in this case.

One value of this thesis is its quantitative approach to operational and combined errors. Such errors are known to exist in any GIS product (Mead, 1982), yet few studies have separated and quantified operational and combined errors to demonstrate the effects which these types of error have on overall GIS product reliability. Specific numerical results of this study could only be directly compared to other areas of similar soils, vegetation, and terrain; as such, these numbers are not easily transferable to many locations outside of the study region. The methodologies used to determine the accuracies of the GIS products, however, can be used to derive GIS accuracy figures for any location in the world. Performing fieldchecks against all base products is admittedly time consuming and costly and therefore not always practical. This study, however, is useful for the accuracy assessment procedures presented and in the proof of its underlying theme that caution should be exercised when using any GIS product; especially when more than two layers are simultaneously examined or when the accuracy of any base product is of unknown or of questionable quality.

Because this study compares maps aggregated by areas to field samples recorded at points, it presents the level of accuracy which might be expected when GIS products aggregated by areas are used to infer point values. The high levels of error which result from such a comparison suggest that special caution be exercised when trying to imply specific point values from products which represent only multilayered generalizations of data previously generalized and aggregated by areas. This idea is similar to the concept described in Vitek, Walsh, and Gregory (1984) that data lost through the process of generalization cannot be recreated from the map product. Implying specific values from generalized areal data will always create error, and any decisions based on such a process are necessarily flawed. Using a GIS to imply only spatial patterns or area trends would be expected to result in less user interpretation error, yet the user who is concerned with such information (for example, measurement of area rather than determination of precise boundary locations) can usually get by with a lower level of accuracy (Marble and Peuquet, 1983).

By controlling (as much as possible) the inherent errors within base products and the operational error created by GIS map production, total or combined error may be minimized. Although the total error thus created in any GIS product will not change once the map is produced, using such a product to make general assumptions about spatial patterns and trends, rather than trying to imply specific point values, will contribute to fewer user interpretation errors and, therefore, more sound GIS-based decisions. If point values are to be assumed from GIS products, the user should at least be aware of the high levels of error possible with any decisions based on such a practice.

A SELECTED BIBLIOGRAPHY

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, "A land use and land cover classification system for use with remote sensor data", U.S. Geological Survey Professional Paper #964, 28p., 1976.

Anderson, K.E. and R.W. Marx, "A prospective case for a national land data system: ten years later", Proceedings, Auto-Carto VII, pp. 1-10, 1985.

Aronoff, S., "Classification accuracy: a user approach", Photogrammetric Engineering and Remote Sensing, 48:8, pp.1299-1307, 1982a.

Aronoff, S., "The map accuracy report: a user's view", Photogrammetric Engineering and Remote Sensing, 48:8, pp. 1309-1312, 1982b.

Averack, R., "Methods and algorithms for boundary definition", Proceedings of the International Symposium on Spatial Data Handling, pp. 252-266, 1984.

Chang, K.T., "Multicomponent quantitative mapping", Cartographic Journal, 19:2, pp.95-103, 1982.

Chrisman, N.R., "Estimating the error of categorical maps", Harvard University Graduate School of Design, 32 p., 1982a.

Chrisman, N.R., "A theory of cartographic error and its measurement in digital data bases", Auto-Carto V Proceedings, pp.159-168, 1982b.

Chrisman, N.R., "The role of quality information in the long-term functioning of a geographic information system", Cartographica, 21:2, pp. 79-87, 1984.

Congalton, R.G., "A comparison of five sampling schemes used in assessing the accuracy of land cover/land maps derived from remotely sensed data".(Unpublished Ph.D. dissertation, Virginia Polytechnic Institute and State University, 1984.)

Crapper, P.F., "An estimate of the number of boundary cells in a mapped landscape coded to grid cells", Photogrammetric Engineering and Remote Sensing, 50:10, pp. 1497-1503, 1984.

Crosley, P., "Creating user friendly geographic information systems through user friendly system supports", Proceedings, Auto-Carto VII, pp. 133-140, 1985.

Elassal, A.A. and V.M. Caruso, "Digital elevation models", Geological Survey Circular 895-B, USGS Digital Cartographic Data Standards, 40 p., 1983.

Estes, J.E., "Remote sensing and geographic information systems coming of age in the 80's", Proceedings,PECORA VII Symposium, pp.23-30, 1982.

Estes, J.E., "Geographic applications of remotely sensed data", Proceedings of the IEEE, 73:6, pp. 1097-1107, 1985.

Franklin, W.R., "Cartographic errors symptomatic of underlying algebra problems", Proceedings of the International Symposium on Spatial Data Handling, pp. 190-208, 1984.

Gates, W.E. and R.J. Heil, "Geographic information systems", Journal of the Surveying and Mapping Division, Proceedings of the American Society of Civil Engineers, 106:SU1, pp.105-117, 1980.

Gersmehl, P.J. and C.A. Gersmehl, "The graphic display of inventory data", Auto-Carto V Proceedings, pp.317-326, 1982.

Ginevan, M.E., "Testing land use map accuracy: another look", Photogrammetric Engineering and Remote Sensing, 45:10, pp.1371-1377, 1979.

Goodchild, M.F., "Statistical aspects of the polygon overlay problem", Harvard Papers on Geographic Information Systems, 6, pp.1-22, 1978.

Greenland, A., R.M. Socher, and M.R. Thompson, "Statistical evaluation of accuracy for digital cartographic data bases", Proceedings, Auto-Carto VII, pp. 212-221, 1985.

Gregory, A.F. and H.D. Moore, "Economical maintenance of a national topographic data base using Landsat images", Photogrammetric Engineering and Remote Sensing, 52:4, pp. 519-524, 1986.

Guptill, S.C., "The impact of computer graphics, data manipulation software, and computing equipment on spatial data structures", Harvard Papers on Geographic Information Systems, 2, pp. 1-9, 1978.

Guptill, S.C., "Interfacing digital cartographic base data with supplementary data in user oriented geographic information systems", Proceedings,International Symposium on Remote Sensing of Environment, pp.377, 1981.

Hansen, D., M. Dwyer, and M. Mogg, "Methods for comparing data entry systems", Proceedings, Auto-Carto VII, pp. 237-243, 1985.

Hart, J.A., D.B. Wherry, and S. Bain, "An operational GIS for Flathead National Forest", Proceedings, Auto-Carto VII, pp. 244-253, 1985.

Hay, A.M., "Sampling designs to test land use map accuracy", Photogrammetric Engineering and Remote Sensing, 45:4, pp.529-533, 1979.

Henderson, F.M., "Effects of interpretation techniques on land use mapping accuracy", Photogrammetric Engineering and Remote Sensing, 46:3, pp.359-367, 1980.

Henderson, J., "The utility of a GIS in evaluating the accuracy of classified Landsat landcover type maps on the Kenai Peninsula", Proceedings, Auto-Carto VII, pp. 257-266, 1985.

Henley, Jim, Personal Interview, Perry Soil Conservation Service, Perry, Oklahoma, December 10, 1985.

Jenks, G.F., "Contemporary statistical maps - evidence of spatial and graphic ignorance", The American Cartographer, 3:1, pp.11-19, 1976.

Jenks, G.F., "Lines, computers, and human frailties", Annals of the Association of American Geographers, 71:1, pp.1-10, 1981.

Kimerling, A.J., "Theoretical and practical relationships between remote sensing and cartography".(Unpublished Ph.D. dissertation, The University of Wisconsin - Madison, 1976.)

Klock, G.O., P. Gum, and L.E. Jordan, "Integrated resource management utilizing multispectral imagery and geographic information systems: Okanogan National Forest", Technical Papers, ACSM - ASPRS Annual Convention, 3, pp. 46-54, 1986.

Lehan, T.J., "The influence of spatial information on data queries", Technical Papers, ACSM - ASPRS Annual Convention, 3, pp. 250-257, 1986.

Marble, D.F., "Geographic information systems - an overview", Proceedings, PECORA IX Symposium, pp. 18-24, 1984.

Marble, D.F. and D.J. Peuquet, "Geographic information systems and remote sensing", Manual of Remote Sensing, 2nd edition, Falls Church,Virginia: American Society of Photogrammetry, pp.923-958, 1983.

Marble, D.F., J.P. Lauzon, and M. McGranaghan, "Development of a conceptual model of the manual digitizing process", Proceedings of the International Symposium on Spatial Data Handling, pp. 146-171, 1984.

Markham, B.L. and J.R.G. Townshend,"Landcover classification accuracy as a function of sensor spatial resolution", Proceedings,International Symposium on Remote Sensing of Environment, pp.1075-1085, 1981.

Matson, J.T., "Applying the SYMAP algorithm for surface compatibility and comparative analysis of areal data", The American Cartographer, 12:2, pp. 113-121, 1985.

McEwen, R.B., H.W. Calkins, and B.S. Ramey, "Overview and USGS activities", Geological Survey Circular 895-A, USGS Digital Cartographic Data Standards, 20 p., 1983.

Mead, D.A., "Assessing data quality in geographic information systems", Remote Sensing for Resource Management, Ed. Chris Johannsen and James Sanders, Iowa: Soil Conservation Society of America, pp. 51-59, 1982.

Monmonier, M.S., "Cartography, mapping, and geographic information", Progress in Human Geography, 7:3, pp.420-428, 1983.

Muller, J.C., "Map gridding and cartographic errors - a recurrent argument", The Canadian Cartographer, 14:2, pp.152-167, 1977.

National Aeronautics and Space Administration and Environmental Systems Research Institute, "Review and synthesis of problems and directions for large scale geographic information system development", Proceedings of the International Symposium on Spatial Data Handling, pp. 19-25, 1984.

National Cartographic Information Center, "Digital terrain tapes - user's guide", 2nd edition, U.S. Geological Survey, National Cartographic Information Center, 12 p., 1979.

Newcomer, J.A. and J. Szajgin, "Accumulation of thematic map errors in digital overlay analysis", The American Cartographer, 11:1, pp.58-62, 1984.

Nichols, J.D., "Characteristics of computerized soil maps", Proceedings, Soil Science Society of America, 39, pp.927-932, 1975.

Peuquet, D.J., "Data structures for a knowledge-based geographic information system", Proceedings of the International Symposium on Spatial Data Handling, pp. 372-391, 1984.

Robinson, G. and M. Jackson, "Expert systems in map design", Proceedings, Auto-Carto VII, pp. 430-439, 1985.

Root, R.R., S.C.F. Stitt, M.O. Nyquist, G.S. Waggoner, and J.K. Agee, "Vegetation and fire fuels model mapping of North Cascades National Park", Technical Papers, ACSM - ASPRS Annual Convention, 3, pp. 46-54, 1986.

Rosenfield, G.H. and K. Fitzpatrick-Lens, "A coefficient of agreement as a measure of thematic classification accuracy", Photogrammetric Engineering and Remote Sensing, 52:2, pp. 223-227, 1986.

Salmen, L.J., "Natural resource analysis techniques - their implications for data base management systems", Harvard Papers on Geographic Information Systems, 2, pp. 1-10, 1978.

Schwarz, D.E., "Variability of the accuracy of delineating agricultural field boundaries from satellite images of the United States".(Unpublished Ph.D. dissertation, University of Kansas, 1976.)

Short, N.M., "The Landsat tutorial workbook: the basics of satellite remote sensing", NASA Reference Publication #1078, 1982.

Sinton, D.F., "The inherent structure of information as a constraint to analysis: mapped thematic data as a case study", Harvard Papers on Geographic Information Systems, 7, pp. 1-17, 1978.

Soil Conservation Service, "Soil surveys nature, purposes, and uses", Soil Survey Manual, Washington, D.C.: Soil Conservation Service: United States Department of Agriculture, pp. 2-1 - 2-17, 1984.

Starr, L.E. and K.E. Anderson, "Some thoughts on cartographic and geographic information systems for the 1980's", Proceedings,PECORA VII Symposium, pp.41-56, 1982.

Story, M. and R.G. Congolton, "Accuracy assessment: a user's perspective", Photogrammetric Engineering and Remote Sensing, 52:3, pp. 397-399, 1986.

Teicholz, E., "Geographic information systems: the ODYSSEY project", Proceedings,American Society of Civil Engineers, Journal of the Surveying and Mapping Division, 106:SU1, pp.119-135, 1980.

Tempfli, K., "Genauigkeitsschatzung digitaler hohenmodelle mittels spektralanalyse (Accuracy estimation of digital elevation models using spectral analysis)".(Unpublished Ph.D. dissertation, Technische Universitat Wien, 1982.)

Traylor, C.T., "The evaluation of a methodology to measure manual digitizing error in cartographic data bases".(Unpublished Ph.D. dissertation, University of Kansas, 1979.)

United States Geological Survey, "Topographic surveys and mapping", U.S Geological Survey Professional Paper #1175, p. 353, 1980.

VanGenderen, J.L., B.F. Lock, and P.A. Vass, "Testing the accuracy of remote sensing land use maps", 11th International Symposium on Remote Sensing of Environment, pp.615-623, 1977.

VanGenderen, J.L., B.F. Lock, and P.A. Vass, "Remote sensing: statistical testing of thematic map accuracy", Remote Sensing of Environment, 7:1, pp.3-14, 1978.

VanHorn, E.K., "Generalizing cartographic data bases", Proceedings, Auto-Carto VII, pp. 532-540, 1985.

Vitek, J.D., S.J. Walsh, and M.S. Gregory, "Accuracy in geographic information systems: an assessment of inherent and operational errors", Proceedings,PECORA IX Symposium, pp.296-302, 1984.

Vitek, J.D. and D.G. Richards, "Incorporating inherent map error into flood hazard analysis", Professional Geographer, 30:2, pp.168-173, 1978.

Vonderohe, A.P. and N.R. Chrisman, "Tests to establish the quality of digital cartographic data: some examples from the Dane County land records project", Proceedings, Auto-Carto VII, pp. 552-559, 1985.

Wallin, E., "Isarithmic maps and geographical disaggregation", Proceedings of the International Symposium on Spatial Data Handling, pp. 209-217, 1984.

Walsh, S.J. "Geographic information systems for natural resource management", Journal of Soil and Water Conservation, 40:2, pp. 202-205, 1985.

Walsh, S.J., J.D. Vitek, and M.S. Gregory, "Landsat digital analysis of forest clearcuts and reforestation", Remote Sensing for Resource Management, Soil Conservation Society of America, pp.159-171, 1982.

Walsh, S.J. and S.J. Stadler, "Regional evapotranspiration modeling of Oklahoma's Little Washita drainage basin", Proceedings, Ninth International Symposium of Machine Processing of Remotely Sensed Data, pp. 389-398, 1983.

Walsh S.J., S.J. Stadler, and M.S. Gregory, "Modeling of regional evapotranspiration through a geographic information system approach", Proceedings of the International Symposium on Spatial Data Handling, pp. 252-266, 1984.

Wehde, M., "Grid cell size in relation to errors in maps and inventories produced by computerized map processing", Photogrammetric Engineering and Remote Sensing, 48:8, pp.1289-1298, 1982.

Wilson, C.L. and F.J. Thomson, "Integration and manipulation of remotely sensed and other data into geographic information systems", Proceedings, PECORA VII Symposium, pp.303-311, 1982.

Wolf-Dieter, R., "Computer controlled input and output of speech: utilizing the acoustic channel for geometric data aquisition", Proceedings of the International Symposium on Spatial Data Handling, pp. 172-180, 1984.

APPENDICES

APPENDIX A

MAP RECORDED DATA FOR 35 SAMPLE POINTS

APPENDIX A

MAP RECORDED DATA FOR 35 SAMPLE POINTS

Landcover Classification from 1:24,000 LANDSAT LANDCOVER MAPS		
SAMPLE POINT #	2.5 acre (100 m^2)	10.0 acre (200 m^2)
1	Sparse Woodland	Forest
2	Sparse Woodland	Sparse Woodland
3	Range	Range
4	Cropland	Cropland
5	Cropland	Range
6	Range	Range
7	Cropland	Range
8	Range	Range
9	Sparse Woodland	Sparse Woodland
10	Exposed Soil	Range
11	Range	Range
12	Sparse Woodland	Sparse Woodland
13	Range	Sparse Woodland
14	Forest	Forest
15	Exposed Soil	Exposed Soil
16	Range	Sparse Woodland
17	Sparse Woodland	Sparse Woodland

APPENDIX A (Continued)

Landcover Classification from 1:24,000
LANDSAT LANDCOVER MAPS

SAMPLE POINT #	2.5 acre (100 m^2)	10.0 acre (200 m^2)
18	Sparse Woodland	Sparse Woodland
19	Cropland	Range
20	Exposed Soil	Range
21	Forest	Forest
22	Sparse Woodland	Water
23	Forest	Sparse Woodland
24	Range	Range
25	Sparse Woodland	Range
26	Range	Exposed Soil
27	Water	Water
28	Cropland	Cropland
29	Range	Forest
30	Forest	Forest
31	Cropland	Cropland
32	Cropland	Cropland
33	Cropland	Cropland
34	Cropland	Cropland
35	Cropland	Cropland

APPENDIX A (Continued)

Slope Angle and Slope Aspect Terrain Data from
1:250,000 USGS DIGITAL TERRAIN TAPE

SAMPLE POINT #	2.5 acre (100 m^2) SLOPE ANGLE %	SLOPE ASPECT [1]	10.0 acre (200 m^2) SLOPE ANGLE %	SLOPE ASPECT [1]
1	4-5	2	1-2	3
2	0-1	6	0-1	5
3	1-2	2	1-2	3
4	0-1	5	0-1	5
5	1-2	3	1-2	4
6	1-2	6	1-2	5
7	1-2	3	1-2	4
8	1-2	3	1-2	3
9	0-1	6	1-2	6
10	1-2	3	1-2	3
11	0-1	6	0-1	6
12	5-6	1	2-3	7
13	5-6	1	2-3	7
14	0-1	1	0-1	6
15	4-5	3	2-3	3
16	1-2	4	1-2	4
17	1-2	5	1-2	4
18	1-2	6	1-2	5
19	1-2	5	1-2	3
20	1-2	3	1-2	4

APPENDIX A (Continued)

Slope Angle and Slope Aspect Terrain Data from
1:250,000 USGS DIGITAL TERRAIN TAPE

SAMPLE POINT #	2.5 acre (100 m^2) SLOPE ANGLE %	SLOPE ASPECT[1]	10.0 acre (200 m^2) SLOPE ANGLE %	SLOPE ASPECT[1]
21	1-2	5	1-2	5
22	1-2	6	1-2	6
23	1-2	6	1-2	5
24	1-2	2	1-2	2
25	1-2	2	1-2	2
26	1-2	5	1-2	5
27	0-1	water	0-1	water
28	1-2	5	0-1	3
29	1-2	2	1-2	2
30	1-2	2	0-1	2
31	1-2	1	1-2	1
32	1-2	1	1-2	1
33	1-2	1	1-2	1
34	1-2	1	1-2	1
35	1-2	1	1-2	1

[1] Slope aspect is given as a number from 1 to 8 corresponding to the 8 major points of the compass as follows:
1 = NORTH, 2 = NORTHEAST, 3 = EAST, 4 = SOUTHEAST,
5 = SOUTH, 6 = SOUTHWEST, 7 = WEST, AND 8 = NORTHWEST.

APPENDIX A (Continued)

SCS Soil Series Classification from 1:24,000 scale (gridded) SCS SOILS MAP[1]					
SAMPLE POINT #	SOIL TYPE AT-A-POINT	CELL MIDPOINT SAMPLE 2.5 acre ($100\ m^2$)	CELL MIDPOINT SAMPLE 10.0 acre ($200\ m^2$)	CELL DOMINANT SAMPLE 2.5 acre ($100\ m^2$)	CELL DOMINANT SAMPLE 10.0 acre ($200\ m^2$)
1	33	33	59	33	59
2	76	11	11	11	11
3	33	35	6	35	35
4	42	6	42	41	42
5	6	42	42	42	42
6	72	72	11	72	11
7	37	6	39	6	42
8	41	37	6	37	37
9	76	76	76	76	76
10	37	37	6	37	6
11	54	2	51	2	2
12	76	76	76	76	76
13	76	76	76	76	76
14	11	11	11	11	11
15	61	35	35	35	35
16	3	3	26	3	72
17	26	25	11	25	61
18	65	65	4	65	4
19	61	35	35	35	35
20	11	3	3	3	3

APPENDIX A (Continued)

		SCS Soil Series Classification from 1:24,000 scale (gridded) SCS SOILS MAP[1]			
SAMPLE POINT #	SOIL TYPE AT-A-POINT	CELL MIDPOINT SAMPLE		CELL DOMINANT SAMPLE	
		2.5 acre (100 m^2)	10.0 acre (200 m^2)	2.5 acre (100 m^2)	10.0 acre (200 m^2)
21	26	water	26	26	26
22	11	25	3	25	11
23	11	11	11	11	11
24	45	26	26	26	26
25	45	26	35	26	26
26	76	76	26	76	26
27	water	water	35	water	35
28	42	42	42	42	42
29	41	35	35	35	35
30	6	41	41	41	35
31	6	42	42	42	42
32	41	41	6	41	6
33	6	41	41	41	41
34	37	37	6	37	6
35	43	43	42	43	42

[1]See APPENDIX C for SCS Soil Series number mapping legend

APPENDIX B

FIELD RECORDED DATA FOR 35 SAMPLE POINTS

APPENDIX B

FIELD RECORDED DATA FOR 35 SAMPLE POINTS

SAMPLE POINT #	LANDCOVER	SLOPE ANGLE %	SLOPE ASPECT[1]	SOILS[2]
1	Range	4	3	33
2	Range	4	3	76
3	Range	2	5	33
4	Cropland	3	3	41
5	Sparse Woodland	16	6	42
6	Range	8	5	72
7	Cropland	6	5	43
8	Cropland	0	2	42
9	Sparse Woodland	13	7	11
10	Cropland	1	6	41
11	Cropland	10	7	54
12	Sparse Woodland	6	6	11
13	Sparse Woodland	8	5	76
14	Forest	3	5	11
15	Cropland	1	8	61
16	Range	9	4	72
17	Range	10	7	61
18	Sparse Woodland	10	7	25
19	Sparse Woodland	9	4	61

APPENDIX B (Continued)

SAMPLE POINT #	LANDCOVER	SLOPE ANGLE %	SLOPE ASPECT[1]	SOILS[2]
20	Bare Soil	4	5	11
21	Forest	39	4	26
22	Sparse Woodland	21	5	11
23	Forest	2	3	11
24	Cropland	5	8	45
25	Range	5	4	3
26	Range	5	6	76
27	Water	0	water	water
28	Cropland	0	1	42
29	Range	1	7	41
30	Forest	6	4	6
31	Forest	28	4	42
32	Forest	35	1	41
33	Cropland	10	4	41
34	Forest	8	7	42
35	Cropland	4	5	not used

[1]Slope aspect is recorded as a number from 1-8 corresponding to 8 major compass points as follows: 1 = N, 2 = NE, 3 = E, 4 = SE, 5 = S, 6 = SW, 7 = W, and 8 = NW.

[2]See APPENDIX C for SCS Soil Series number mapping legend.

APPENDIX C

SCS SOILS MAPPING LEGEND

APPENDIX C

SCS SOILS MAPPING LEGEND

SOIL CORRELATION OF PAYNE COUNTY, OKLAHOMA
SOIL SURVEY, 1984

Publication Symbol	Approved Mapping Unit Name
2	Coyle loam, 1 to 3 percent slopes
3	Coyle loam, 3 to 5 percent slopes
4	Coyle loam, 2 to 5 percent slopes, eroded
6	Pulaski fine sandy loam, frequently flooded
11	Stephenville-Darnell complex, 1 to 8 percent slopes
25	Grainola-Lucien complex, 1 to 5 percent slopes
26	Grainola-Lucien complex, 5 to 12 percent slopes
33	Norge loam, 1 to 3 percent slopes
35	Norge loam, 2 to 5 percent slopes, eroded
37	Port silt loam, occasionally flooded
39	Port-Oscar complex, occasionally flooded
41	Easpur loam, occasionally flooded
42	Ashport silty clay loam, occasionally flooded
43	Pulaski fine sandy loam, occasionally flooded
45	Renfrow silt loam, 1 to 3 percent slopes

APPENDIX C (Continued)

SOIL CORRELATION OF PAYNE COUNTY, OKLAHOMA
SOIL SURVEY, 1984

Publication Symbol	Approved Mapping Unit Name
51	Stephenville fine sandy loam, 1 to 5 percent slopes, severely eroded
54	Stephenville fine sandy loam, 3 to 5 percent slopes
59	Konawa and Teller soils, 2 to 6 percent slopes, eroded
61	Mulhall loam, 3 to 5 percent slopes, eroded
65	Grainola clay loam, 3 to 5 percent slopes
72	Zaneis-Huska complex, 1 to 5 percent slopes
76	Coyle and Zaneis soils, 2 to 5 percent slopes, severely eroded

VITA

Dale Robert Lightfoot

Candidate for the Degree of

Master of Science

Thesis: ASSESSMENT OF INHERENT AND OPERATIONAL ERRORS IN GEOGRAPHIC INFORMATION SYSTEMS

Major Field: Geography

Biographical:

Personal Data: Born in Fort Worth, Texas, November 10, 1957, the son of Frank K. and Loretta Lightfoot. Married to Victoria K. Harris on July 28, 1979.

Education: Graduated from Central High School, Tulsa, Oklahoma, in May, 1976; received Bachelor of Science degree in Industrial Arts Education from Oklahoma State University in May, 1980; completed requirements for the Master of Science degree at Oklahoma State University in July 1986.

Professional Experience: Research Assistant, Department of Geography, Oklahoma State University, January 1985 to May 1986; Teaching Assistant, Department of Geography, Oklahoma State University, August 1985 to May 1986.